Towards Predicting Individual Treatment Response in Patients with Multiple Sclerosis

Dr Sarmad Adnan Hameed Al-Araji [redacted]

Queen Square MS Centre

Department of Neuroinflammation

UCL Institute of Neurology

London WC1N 3BG

Primary Supervisor: Professor Olga Ciccarelli

Subsidiary Supervisor: Dr Ashwani Jha

Tertiary Supervisor: Professor Claudia Wheeler-Kingshott

Thesis Committee: Professor Parashkev Nachev, Dr Baris Kanber

Thesis submitted to University College London for the degree of Doctor of Philosophy October 2022

Disclaimer

I, Sarmad Adnan Hameed Al-Araji confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Dr Le Zhang and Dr Baris Kanber assisted with the imaging acquisition and analysis presented in chapters 4 and 5. Dr Ashwani Jha, secondary supervisor, Dr Marcello Moccia, Assistant Professor at the University of Naples Federico II and Clinical Research Fellow at University College London, Dr Raffaele Palladino, clinical epidemiologist and statistician at school of public health, Imperial college London and Professor Maria Pia Sormani, Professor of biostatistics at the University of Genoa, Italy assisted with the statistical analysis presented in chapters 3, 4 and 5.

Abstract

This thesis demonstrates an observational predominantly retrospective study of nearly 2000 consecutive relapsing remitting multiple sclerosis (RRMS) patients recruited after initiating disease modifying therapies (DMTs); glatiramer acetate, dimethyl fumarate, fingolimod, natalizumab or ocrelizumab, in the National Health Service at the National Hospital for Neurology and Neurosurgery and followed up for a period of two years.

Due to the heterogeneity of onset, diagnosis, classification, treatment response and prognosis, it is difficult to predict outcomes in MS patients. Machine learning has been evolving in the medical field over the last two decades and we have therefore reviewed the literature on the use of machine learning (ML) in MS diagnosis, classification, progression and prediction. Despite the enormous application on ML in the field of MS in the last 10 years, we found limited studies that have focused on predicting the response to DMTs that could assist clinicians in choosing the best treatment for MS patients. We therefore explored the association of routinely available clinical and radiological variables in predicting NEDA (no evidence of disease activity) at the group-level in our separate cohorts. We found that baseline variables including older age at MS onset, lower number of previous DMTs, lower number of relapses in the previous 12 months, lower number of new and/or Gd-enhancing MRI brain lesions, and lower EDSS score showed significant association of achieving NEDA in most cohorts.

Whilst these factors are associated with clinical response at the group-level, we are unable to make clinically useful predictions of treatment response at the individual-level. We therefore built a Bayesian modelling framework optimal for estimating individual-level uncertainty in predicting NEDA. We combined all cohorts and built five predicting models with increasing complexity. The most comprehensive model with 12 clinical and radiological predictors showed the best accuracy, reaching 73% and 69%, in predicting individualised NEDA at 1 and 2 years, respectively.

Finally, since there have been no head-to-head phase III clinical trials comparing multiple DMTs with variable efficacy, we conducted a pilot study to compare the clinical and radiological effectiveness of five different DMTs in our real-world cohort using propensity score adjustment. Although ocrelizumab showed slightly better outcomes, considering the limitations of this study, including the observational and retrospective nature of the study, different therapies used over two decades and the small sample size, further larger studies and using propensity score matching are needed in the absence of randomised controlled trials.

Impact statement

Benefits inside academia

Large real-world data is notoriously difficult to collect and tends to be time consuming. Here, I have shown that a large dataset of clinical and radiological variables, collected by myself, has been achieved from a single tertiary MS centre, which has shown evidence of significant associations with treatment outcomes. The interpretation of the results should still be cautious considering the observational nature of the study including the presence of selection bias and confounding factors, and the need for validation in other external independent cohorts. The presentation of the preliminary findings of this project received interest from the multiple sclerosis scientific community and it was presented initially at the annual ECTRIMS meeting in October 2021. Further results were presented at the ABN conference in May 2022.

My approach in building a Bayesian modelling framework optimal for estimating individual-level uncertainty in predicting treatment outcomes has not been done before in the field of MS and has attracted scientific interest. This work was presented at the annual ECTRIMS meeting in October 2022 and won the prize for the best poster. In the meantime, I am collaborating with the Swedish MS team to validate our models in an unseen Swedish MS population.

Creating this invaluable database of nearly 2000 patients will undoubtedly be an asset that can be explored further during my post-doctoral fellowship applications and will be helpful for other scientists wishing to answer a particular question or a hypothesis. Some of this data has already been used in supporting other researchers as part of the Magnetic Resonance Imaging in Multiple Sclerosis (MAGNIMS), a consortium that has made a significant contribution to defining the role of MRI in diagnosis and monitoring treatments in MS, looking at the response to oral therapies in MS. The preliminary results were presented at ECTRIMS October 2021 and further results were presented at ECTRIMS October 2022.

Benefits outside academia

The results presented in my thesis have potential impacts in the clinical setting. The vast majority of clinical and radiological characteristics demonstrated in this thesis are usually readily available to the MS treating clinician, not just in the NHS, world-wide. My findings can help in understanding the important baseline clinical and radiological characteristics that may impact on the likeliness of response to a particular therapy, not just at the group-level, but also at the individual-level. This work has created the basis for a future provision to support clinicians and patients when a treatment is needed for new RRMS patients and those who fail therapy and need to switch. Further validation in other external large cohorts and larger studies are needed in the absence of randomised controlled trials.

Finally, with the limitation of an observational study which provides a lower level of evidence compared to systematic reviews and randomised control trials, I provided class III information for evaluating the effectiveness of multiple DMTs in the management of RRMS. Patients taking lower, moderate and higher efficacy DMTs continue to be at risk of experiencing disease activity in the first two years. Although we provided some indications on how effective DMTs are in our cohorts, further long-term studies including registry-based trials evaluating more DMTs in larger cohorts are needed to compare different therapies and investigate risk factors for patients experiencing disease activity.

Acknowledgments

I have had the pleasure to work with many colleagues at Queen Square MS Centre in the past three years whom without this work would not have been possible. I would like to thank:

- My primary supervisor, Professor Olga Ciccarelli, for giving me the opportunity to work within her team. I have learned a lot from her expertise and knowledge both clinically and academically. I am grateful for the support and mentorship which helped me immensely during my research.
- My secondary supervisor, Dr Ashwani Jha, for his support and guidance. I am thankful for his help in understanding and learning academic work and statistical analysis, particularly the use of R. His continuous support in guiding my development and progress has been crucial in helping me structure my work.
- My thesis committee, Professor Parashkev Nachev, Dr Baris Kanber and Le Zhang for their guidance, support, and helpful feedback.
- My colleagues at Queen Square MS Centre including all the MS consultants (Dr Wallace Brownlee, Dr Declan Chard, Prof Jeremy Chataway, Dr Karen Chung, Dr Floriana De Angelis, Dr Zhaleh Khaleeli, Dr Siobhan Leary, Dr Jo Swanton, Prof Alan Thompson, Prof Ahmed Toosy, Dr Anand Trip and Dr Heather Wilson) and the PITMS study team; Dr Alessia Bianchi, Dr Charmaine Yam, Dr Omar Abdel-Mannan, Dr Olivia Goodkin, Dr Arman Eshaghi, Miss Amber Strang, Miss Eirini Samdanidou, Mr Alvin Zapata, Dr Anuriti Aojula, Dr Dominic Wilkins, Dr Yael Hacohen and Professor Frederik Barkhof for the ongoing help and support particularly for the prospective PITMS study for which recruitment and assessments are ongoing.

I would like to thank my parents, Adnan and Wafaa, for their unconditional life-long love and support.

And finally, I would like to thank Charlotte (my wife), Ramsey (my son) and Raya (my daughter) who have helped me to deal with the challenges throughout my clinical and academic career. I have been truly blessed to have their support which I will always be grateful for.

Presentations associated with this thesis

The following presentations were produced using work done as part of this project:

1. **Al-Araji S**, Jha A, Zhang L, Eshaghi A, Kanber B, Bianchi A, Yam C, Abdel-Mannan O, Aojula A, Champsas D, Goodkin O, Pontillo G, Hamed W, Wilkins D, Brownlee W, Chard D, Chataway J, Chung K, De Angelis F, Khaleeli A, Leary S, Swanton J, Thompson A, Toosy S, Trip A, Wilson H, Barkhof F, Nachev P, Ciccarelli O. Bayesian prediction of individualised treatment response in multiple sclerosis. Poster presented at ECTRIMS October 2022 and won the best poster award.
2. Moccia M, **Al-Araji S**, Zhang L, Abdel-Mannan O, Aojula A, Bianchi A, Champsas D, Eshaghi A, Goodkin O, Hamed W, Pontillo G, Yam C, Wilkins D, Brownlee W, Chard D, Chataway J, Chung K, De Angelis F, Khaleeli Z, Leary S, Swanton J, Thompson A, Toosy AT, Trip SA, Wilson H, Barkhof F, Ciccarelli O. Comparing clinical and radiological effectiveness of disease modifying treatments in the real-world. Presented at ECTRIMS October 2022.
3. Ruggieri S, Prosperini L, **Al-Araji S**, Annovazzi P, Bisecco A, Ciccarelli O, De Stefano N, Filippi M, Fleischer V, Evangelou N, Enzinger C, Gallo A, Garjani A, Groppa S, Haggiag S, Khalil M, Lucchini M, Mirabella M, Montalban X, Pozzilli C, Preziosa P, Rio J, Rocca MA, Rovira A, Stromillo ML, Zaffaroni M, Tortorella C, Gasperini C, MAGNIMS Study Group. Assessing treatment response to oral drugs for Multiple Sclerosis in real world setting: a MAGNIMS study. Abstract and Poster presented at ECTRIMS October 2022.
4. **Al-Araji S**, Bianchi A, Eshaghi A, Zhang L, Kanber B, Jha A, Goodkin O, Barkhof F, Nachev P, Ciccarelli O. Assessing predictors of NEDA in RRMS patients initiating dimethyl fumarate in a real-world setting. Abstract and Poster presented at the ABN May 2022. (final results after 24 months)
5. **Al-Araji S**, Bianchi A, Eshaghi A, Zhang L, Kanber B, Jha A, Goodkin O, Barkhof F, Nachev P, Ciccarelli O. Assessing predictors of no-evidence of disease activity in RRMS patients initiating Dimethyl Fumarate in a real-world setting. Abstract and e-Poster presented at ECTRIMS October 2021. (preliminary results after 12 months)

List of Tables

List of Figures

List of abbreviations:

Abbreviation	Definition
ACE	Angiotensin Converting Enzyme
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARR	Annual Relapse Rate
AUC	Area Under the receiver-operator Curve
BBB	Blood Brain Barrier
BVMT	Brief Visuospatial Memory Test
CIS	Clinically Isolated Syndrome
CMSC	Consortium of MS Centres
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSF	Cerebral Spinal Fluid
CVLT	California Verbal Learning Test
DIS	Dissemination in Space
DIT	Dissemination in Time
DMF	Dimethyl Fumarate
DMT	Disease Modifying Therapy
DSS	Disease Severity Scale
DTI	Diffusion Tensor Imaging
ECG	Electrocardiogram
EDSS	Expanded Disability Status Scale
EEG	Electroencephalogram
ELPD	Expected log posterior density
EOPMS	Early Onset Paediatric MS
EPIC	Electronic Patient Medical Record System
ERP	Event Related Potentials
FBC	Full Blood Count
FDO	First Dose Observations
FLAIR	Fluid Attenuation Inversion Recovery
FMRI	Functional Magnetic Resonance Imaging
GCA	Granger Causality Analysis
GCL	Ganglion Cell Layer
Gd	Gadolinium
HC	Healthy Control
HCL	Hierarchical Clustering
HLA	Human Leukocyte Antigen
HPV	Human Papillomavirus
HR	High Responders
IFN-β	Interferon Beta

IIH	Idiopathic Intracranial Hypertension
IMID	Immune Mediated Inflammatory Diseases
JCV	John Cunningham Virus
LFT	Liver Function Test
LOO	Leave-One-Out
LOO-CR	Leave-One-Out Cross-Validation
LOPMS	Late Onset Paediatric MS
LR	Low Responders
MAGNIMS	Magnetic Resonance Imaging in MS
MCMC	Markov Chain Monte Carlo
MEP	Motor Evoked Potential
MFIS	Modified Fatigue Impact Scale
MICE	Multiple Imputation by Chained Equation
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSIS	Multiple Sclerosis Impact Scale
MSQOL	MS Quality of Life
MSSS	MS Severity Score
NAA	N-acetyl-aspartate
NAIMS	North American Imaging in MS
NAWM	Normal Appearing White Matter
NEDA	No Evidence of Disease Activity
NHNN	National Hospital for Neurology and Neurosurgery
NHS	National Health Service
NIHR	National Institute for Health and Research
NMO	Neuromyelitis Optic
OCB	Oligoclonal Bands
OCT	Optical Coherence Tomography
ON	Optic Neuritis
OR	Odd Ratio
PCA	Principle Component Analysis
PITMS	Predicting Individual Treatment responses towards personalised medicine in Multiple Sclerosis
PML	Progressive Multifocal Leukoencephalopathy
PMS	Progressive MS
PPD	Purified Protein Derivative
PPMS	Primary Progressive MS
QSMSC	Queen Square Multiple Sclerosis Centre
RIS	Radiologically Isolated Syndrome
RNA	Ribonucleic Acid
RNFL	Retinal Nerve Fibre Layer
RRMS	Relapsing Remitting Multiple Sclerosis

SDMT	Symbol Digital Modalities Test
SE	Standard Error
SNP	Single Nucleotide Polymorphism
SPM	Statistical Parametric Mapping
SPMS	Secondary Progressive Multiple Sclerosis
STIR	Short Tau Inversion Recovery
SVM	Support Vector Machine
TB	Tuberculosis
TFT	Thyroid Function Test
TM	Transverse Myelitis
TNF	Tumour Necrosis Factor
UC	Ulcerative Colitis
UE	Urea and Electrolyte
UK	United Kingdom
UTI	Urinary Tract Infection
VZV	Varicella Zoster Virus

Table of Contents

Chapter 1. Multiple Sclerosis

1.1 Introduction:

Multiple Sclerosis (MS) is a chronic inflammatory condition of the central nervous system (CNS) that causes disability with major impact on physical ability, cognition and employment. (1) There are more than 2.8 million MS patients worldwide (2) and, according to the MS society and Mackenzie et al. more than 130,000 people in the United Kingdom (UK) have MS. (3) MS has three main clinical types; relapsing remitting MS (RRMS), primary progressive MS (PPMS) and secondary progressive MS (SPMS). (4) The majority of people with MS (pwMS) present with RRMS (85%) and the rest (15%) present with PPMS. (1, 5, 6)

In about 10-20% of cases, RRMS patients develop a phase of insidious irreversible worsening of neurological function independent of relapses that is called SPMS. (7-9) This is much lower than what was previously reported, (10-12) and this could reflect the changes in the natural history of the disease and the effect of disease modifying therapies. The time from the first clinical presentation of MS to the onset of SPMS is variable, however the median time is approximately 20 years in untreated patients. (13, 14) In a large cohort of RRMS patients, where 62% were treated with DMTs, only 11.3% transitioned to SPMS after a 17-year period. (15)

The clinical phenotypes were revised in 2014 to include disease activity, both clinical and radiological, and disease progression. (16) This was important to improve communication, prognostication, design and recruitment of clinical trials, and therapy decisions. This is summarised in Figure 1: Multiple Sclerosis phenotype description of relapsing and progressive disease.

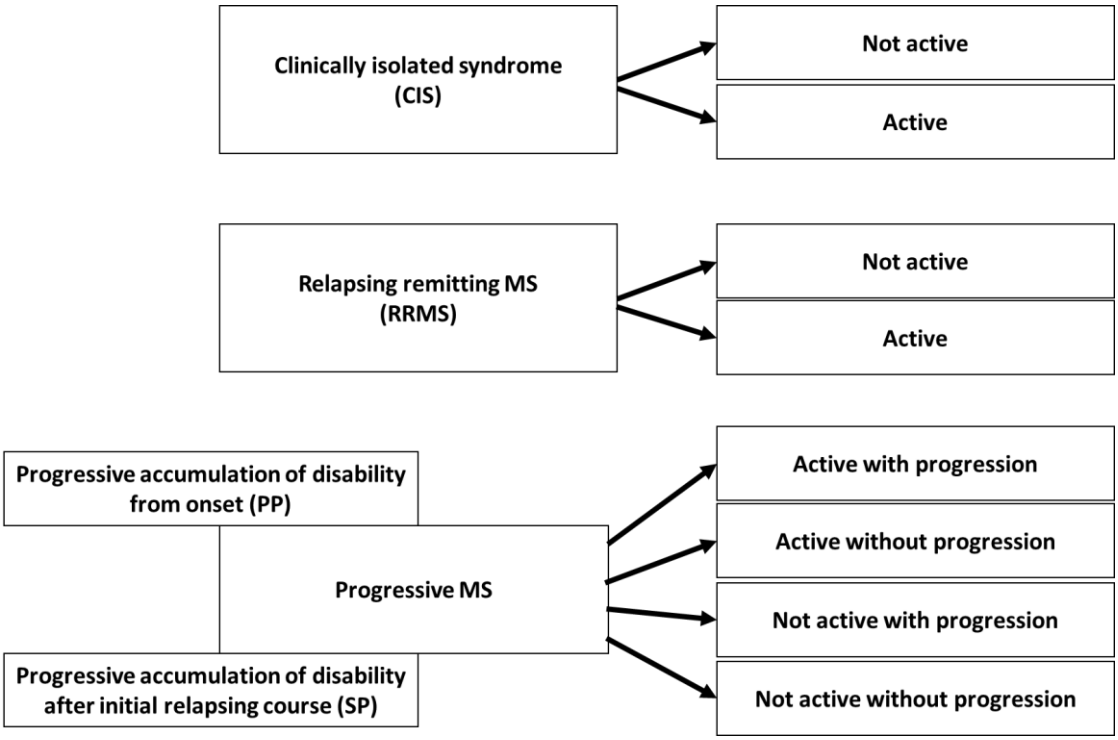


Figure 1: Multiple Sclerosis phenotype description of relapsing and progressive disease.

Activity is determined by clinical relapses assessed at least annually and/or MRI activity (contrast-enhancing lesions; new and enlarging T2 lesions). Progression is measured by clinical evaluation, assessed at least annually.

1.2 Causes of MS

Despite the extensive research into the risk factors that cause MS, the exact cause remains unknown. Multiple risk factors have shown to play a role in developing MS; both environmental and genetics. Many environmental factors have been researched and found to affect the incidence of MS, including Infectious Mononucleosis (Epstein-Barr virus infection) (17), vitamin D level (18, 19), smoking (20, 21), and latitude (22).

The lifetime risk of developing MS in the general population is estimated to be 0.2% and sibling of affected people have up to 20-fold higher risk of developing MS (2-4%), but this risk drops down to 1% in second degree relatives. (1, 6, 23) The concordance rate in monozygotic twins is significantly higher, 30%. (24) In the 1970s, the association between alleles of the major histocompatibility complex and MS was discovered, including DR15, DQ6, and the DRB1*1501, DRB5*0101, DQA1*0102, and DQB2*0602 genotypes. (25, 26) HLADRB1*17 (less pathogenic than HLA-DRB1*15) and HLA-DRB1*08 (more than doubles the risk associated with a single copy of HLA-DRB1*15) were also found to increase the risk of MS. Over the last decade, more alleles have been found as inherited risk factor of MS including alleles of interleukin-2 receptor alpha gene (IL2RA) and interleukin-7 (IL7RA). (27) On the other hand, the following have been suggested to be protective against MS: HLADRB1*14, HLADRB1*01, and HLADRB1*10. (28)

1.3 Pathology of MS

The hallmark of MS pathology are demyelinating plaques. According to autopsy studies these plaques are found in all types of MS. Loss of myelin and reduction of oligodendrocytes, which are responsible for myelin sheaths production and maintenance, are seen in MS plaques. (29, 30) Active inflammatory white matter lesions are the hallmark in RRMS, whereas cortical demyelination and diffuse injury in the normal appearing white matter dominate progressive MS. (30, 31)

The pathological process in MS starts with the loss of regulatory pathways across the blood-brain barrier, causing autoreactive lymphocytes to migrate across, leading to an immune response in the brain. Failure of local regulatory mechanisms with infiltration of perivascular CD8+ T-cells were found to be responsible in developing the MS plaques. (1) Subtypes of T-cells secreting interleukin-17 drive the inflammatory reaction. T-helper 17 cells penetrate the blood-brain barrier when damaged by interleukin 17 and 22, causing death of neurons. (32, 33) The immune response is amplified by the accumulation of T and B cells (lymphocytes), macrophages, plasma cells and pro-inflammatory cytokines recruiting naïve microglia. A deadly signal by cell surface bound tumour necrosis factor α is activated when the activated microglia and oligodendrocyte-myelin unit come together, causing extensive axonal injury typically seen in the acute demyelinating plaques. (29, 34)

Cortical involvement with increasing lesions volume and cortical atrophy is seen in all types of MS, and more extensively in PPMS. Meningeal inflammation is also found in all types of MS, but the most structured form, lymphoid follicles, has not been found in PPMS. (1, 6, 35, 36) This could potentially explain progression clinically despite reduced cerebral inflammation radiologically. (30)

1.4 Pathophysiology of MS

The mechanisms by which MS symptoms arise secondary to the pathological processes are partially understood. Investigators have used data from animal models of MS, both in vitro and in vivo, pathological and magnetic resonance imaging (MRI) data in humans to hypothesise the mechanisms that lead to relapses, intermittent symptoms and clinical progression.

1.4.1 Relapse-related neurological impairment

Inflammation, demyelination and neuroaxonal transection are processes seen in the acute MS lesions which lead to neurological deficit during a relapse. (37, 38)

The healthy myelin sheath in the white matter acts as insulation that helps the action potential to propagate rapidly. However, acutely demyelinated internodes, which have low level of voltage gated sodium channels, cause a conduction block in the affected axons. This is because the nodal sodium currents being inadequate to depolarise the demyelinated axon, causing breakdown in the associated pathways. (39, 40)

1.4.2 Regaining neurological function

During remission (recovery of neurological function), despite restoration of conduction as the acute inflammation settles with the reduction of oedema and damaging inflammatory products, (41) there appears to be ongoing delay in the electrophysiological conduction. (39) Immunological studies showed that Voltage gated sodium channels in the demyelinated axons may reallocate and proliferate leading to enhancement in propagation of the action potential. (42) Partial remyelination plays a part in restoration of normal axonal conduction in the new MS lesions. (43) In addition, previous studies using functional MRI investigating cortical reorganisation after optic neuritis have demonstrated dynamic processes within the occipital cortex during recovery suggestive of adaptive brain plasticity. (44)

1.4.3 Paroxysmal neurological symptoms

Intermittent positive neurological symptoms in MS are thought to be caused by ectopic activity and/or mechanosensitive discharges at the affected axons. (45) Tonic spasms and neurogenic pain are examples of paroxysmal ectopic activity. Visual phosphenes and Lhermitte's phenomenon are examples of mechanosensitive activities. (46)

Temporary conduction block in the previously impaired pathways cause intermittent neurological symptoms. For example, Uhthoff's phenomenon, which is neurological symptoms experienced in heat, is caused by the malfunction of sodium channel opening caused by increase in body temperature and leading to failure of membrane depolarisation. (47)

Exercise induced symptoms and signs are linked to the sustained electrical activity inducing membrane hyperpolarisation via the induction of the electrogenic Na/K ATPase pump causing depolarisation dysfunction. (40, 48)

1.4.4 Disability progression

The progression of neurological deficits arises from the accumulation of focal demyelinating lesions and the gradual neuroaxonal loss over time. (49)

Pathological studies in progressive MS have shown diffuse axonal loss in the normal appearing white matter, axonal transection in focal lesions, and generalised brain atrophy. (31, 37, 50) The extent of brain atrophy in progressive MS is much greater compared to the atrophy seen at the initial stages of MS. (31) This tends to correlate with the extent of neurological impairment and disability. (51) As well as the acute lesions, which contain damaged axons that degenerate overtime, the chronically demyelinated axons degenerate overtime too, leading to progressive neurological disability. (38)

1.5 Measuring neurological impairment

Quantifying MS related neurological deficit based on function, disability and quality of life have led to the development of a variety of clinical scales which have been widely used in MS research. It is essential to be familiar with these scales to understand the studies being published and interpret the data accurately as most of the scales have their own limitation.

The Expanded Disability Status Scale (EDSS) is the most commonly used parameter for the assessment of neurological impairment. (52) During the assessment of EDSS, the clinician assesses the following functional systems; 1. visual, 2. brainstem, 3. pyramidal, 4. cerebellar, 5. sensory, 6. bowel/bladder, 7. cerebral and 8. ambulatory function. A total score of 0-10 (0 = no detectable neurological deficit and 10 = death secondary to MS) is given depending on the scores given for each functional system, as demonstrated in Table 1: EDSS Score. It is important to note that EDSS is an ordinal scale and not normally distributed as transitions between levels are not equal, the lower values of impairments (0-4.0) are based on neurological examination and the higher values (≥4.5) are heavily reliant on walking abilities and loss of independence.

EDSS is widely used as it is easily performed and intuitively interpreted by any clinician familiar with the standard neurological examination. It is therefore the main scale used in this thesis. However, it is important to note that one of the main limitations of EDSS is that it is regarded as heavily weighted on mobility and not very sensitive to cognitive impairment or upper limb dysfunction. (53)

The following scales have also been used widely in MS research; SDMT (Symbol Digital Modalities Test), BVMT-R (Brief Visuospatial Memory Test - Revised) and CVLT-II (California Verbal Learning Test – Edition II). Patient reported outcome measures (PROMS) are validated outcome measures that describe the patient's health status, function and experience through direct reporting. PROMS have been used increasingly in MS clinical trials and practice as they are important for exploring the effects that MS and DMTs have on patients' lives. However, there are some limitations to the use of PROMS as primary outcome in clinical trials due to the lack

of a set of standard measures. (54) PROMS that have been used in MS research including; MS Quality of Life (MSQOL-54), Multiple Sclerosis Impact Scale (MSIS29), Eq5d5I, Modified Fatigue Impact Scale (MFIS) and the Work Productivity and Activity Impairment Question. These scales have not been discussed further as they are not routinely included in clinical practice and have not been included in this thesis.

Table 1: EDSS Score	
Score	**Description**
0	Normal neurological exam, no disability in any functional score
1	No disability, minimal signs in one functional score
1.5	No disability, minimal signs in more than one functional score
2	Minimal Disability in one functional score
2.5	Mild disability in one functional score or minimal disability in two functional scores
3	Moderate disability in one FS, or mild disability in three or four functional scores. No impairment to walking
3.5	Moderate disability in one FS and more than minimal disability in several others. No impairment to walking
4	Significant disability but self-sufficient and up and about some 12 hours a day. Able to walk without aid or rest for 500 metres
4.5	Significant disability but 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. Able to walk without aid or rest for 300m
5	Disability severe enough to impair full daily activities and ability to work a full day without special provisions. Able to walk without aid or rest for 200 metres
5.5	Disability severe enough to preclude full daily activities. Able to walk without aid or rest for 100 metres
6	Requires a walking aid – cane, crutch, etc. – to walk about 100 metres with or without resting
6.5	Requires two walking aids – pair of canes, crutches, etc. – to walk about 20 metres without resting
7	Unable to walk beyond approximately 5 metres even with aid. Essentially restricted to wheelchair; though wheels self in standard wheelchair and transfers alone. Up and about in wheelchair some 12 hours a day
7.5	Unable to take more than a few steps. Restricted to wheelchair and may need aid in transferring. Can wheel self but cannot carry on in standard wheelchair for a full day and may require a motorised wheelchair
8	Essentially restricted to bed or chair or pushed in wheelchair. May be out of bed itself much of the day. Retains many self-care functions. Generally, has effective use of arms
8.5	Essentially restricted to bed much of day. Has some effective use of arms retains some self-care functions
9	Confined to bed. Can still communicate and eat
9.5	Confined to bed and totally dependent. Unable to communicate effectively or eat/swallow
10	Death due to MS

1.6 MRI in MS

MRI has provided critical information for researchers in detecting and quantifying pathology leading to further understanding of the natural history of MS. The introduction of MRI in clinical practice for more than three decades has deepened our knowledge regarding the natural history and pathological processes that lead to neurological impairment in MS. Since it is safe and radiation free, it has been an integral part and the gold standard imaging technique in the diagnosis, longitudinal monitoring and evaluation of therapeutic response in MS. Despite this conventional MRI scans have low specificity in distinguishing between the different diseases and their evolution. This has since directed researchers to develop various techniques and modalities to recognise the different disease processes and therefore led to the production of new therapies. (55)

In the field of MS, the most common MRI sequences used in clinical practice are: T1-weighted (T1w), T2-weighted (T2w), fluid attenuation inversion recovery (FLAIR) and contrast enhancing T1w imaging.

1.6.1 T1-weighted Imaging

T1w imaging show that MS lesions have different intensities with hypo-intense lesions, otherwise known as black holes, indicating more destructive lesions with greater myelin and axonal loss. (38, 56) Higher number of black holes tend to be associated with increased MS progression and disability. (57) This is demonstrated in Figure 2: MRI sequences commonly used in MS

This modality has been shown to be challenging when used without contrast in assessing lesion load due to the similarity in the intensity between lesions and normal appearing white matter (NAWM). Studies of the correlation between clinical outcomes (cognitive outcomes and physical disability) and lesion volume measurement in this modality has been inconsistent with more evidence toward lack of correlation. (58)

1.6.2 T2-weighted Imaging

T2w imaging is sensitive for detecting white matter lesions (Figure 2: MRI sequences commonly used in MS) which are crucial for the diagnosis of RRMS and PPMS. (59) They are more sensitive in detecting MS lesions that are not visible on T1w sequences. However, the appearance of the lesions seen in this modality are not specific and cannot differentiate between the MS pathological phases such as demyelination, axonal loss or remyelination. (60)

T2w sequences are commonly used in MS to measure total brain lesion volume to assess the extent of disease burden. They are also used to assess the conversion of clinically isolated syndrome (CIS) to MS, inflammatory activity, disease progression and response to medication particularly when serial scans are performed at different time points. However, the clinical correlation between disability scores and brain T2 lesion load has not been high as it does not take into account the anatomical location of lesions in the brain, whereas T2 lesions in the spinal cord have been highly associated with MS disability. (58)

1.6.3 Fluid attenuated inversion-recovery

Fluid attenuated inversion-recovery (FLAIR) is a T2w modality which demonstrates subtle periventricular white matter lesions more clearly in relation to dark cerebral spinal fluid (CSF) and hypointense white matter (Figure 2: MRI sequences commonly used in MS). In this sequence MS lesions tend to have the highest intensity followed by grey matter, white matter and CSF. The relative sensitivity of FLAIR images for detecting pathologically confirmed white matter lesions is about 71% at 1.5T MRIs, which is higher compared to the sensitivity of T2w for detecting lesions, 63%. (61) Spinal cord lesions can also be detected using this modality and differentiating plaques with equivocal signal intensity on T2w sequence. (62)

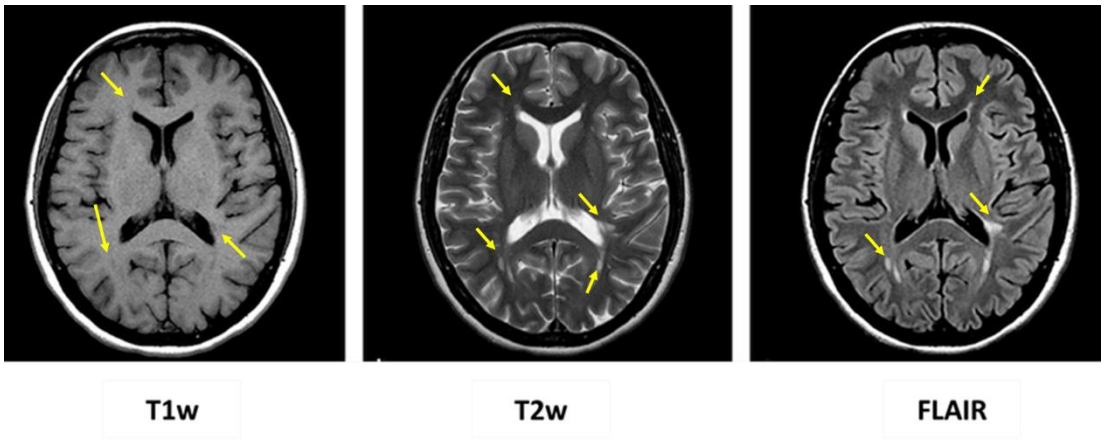


Figure 2: MRI sequences commonly used in MS.
These are the typical sequences used in routine clinical practice. The images have been obtained from an RRMS patient who is taking dimethyl fumarate.

1.6.4 Contrast enhanced T1-weighted Imaging

Contrast enhanced T1w imaging has been used to detect and monitor MS activity. The contrast agent, Gadolinium (Gd), crosses the blood-brain-barrier (BBB) and, during T1 relaxation, enhances active lesions. (41) Gd-enhanced T1w imaging has been central to MS diagnosis, particularly in confirming dissemination in time after one clinical attack or CIS. (63) MS lesions that show Gd enhancement have also been used in assessing the effectiveness of DMTs in both RRMS and in SPMS. (64) This modality is also commonly used to look for MS mimics, for example cerebral sarcoidosis and progressive multifocal leukoencephalopathy (PML), which can be a consequence of MS DMTs such as Natalizumab. (65, 66)

Although Gd is considered largely safe in people with normal kidney function, there remain to be low risk of possible side effects and allergic reactions. The reported frequency of all adverse events after Gd administration range from 0.07-2.4%, severe allergic reactions 0.004-0.7% and life-threatening anaphylactic reactions 0.001-0.01%. (67) Also, considering the concern regarding Gd deposition in brain the European Medicines Agency suspended the use of linear Gd based contrast agent for CNS MRI examinations and recommended the use of the lowest Gd dose only when essential. The recent guidance on MRI and Gd use published in the lancet neurology will be discussed in the next section. (68)

Many other modalities have been used in the field of MS but not necessarily in routine clinical practice and are therefore not discussed in more details in this thesis.

1.6.5 MRI Guidance for Clinicians

In the most recent international consensus, guidelines on how and when to use MRI for diagnosis, prognosis, and treatment monitoring of MS were updated and published by a collective effort from the following groups; Magnetic Resonance Imaging in MS (MAGNIMS), Consortium of MS Centres (CMSC) and North American Imaging in MS (NAIMS). (68) This focused on the use of standardised MRI protocols and the use of Gadolinium. This is summarised in

Table 2: The 2021 MAGNIMS-CMSC-NAIMS consensus on the use of MRI in MS (68)

.

Table 2: The 2021 MAGNIMS-CMSC-NAIMS consensus on the use of MRI in MS (68)

MRI in MS diagnosis:

- **Initial MRI brain protocol:**
  - At least 1.5 Tesla (magnetic field strength)
  - Core sequences include T2w 3D-FLAIR, axial T2w, and T1w with Gadolinium contrast

- **Initial spinal cord MRI protocol:**
  - At least 1.5T MRI
  - At least two of: sagittal T2w sequences, proton density-weighted, or short tau inversion recovery (STIR)
  - Sagittal T1w sequences post Gadolinium contrast

- **Follow up MRI if diagnostic criteria not fulfilled:**
  - MRI brain every 6-12 months in clinically isolated syndrome (CIS) and radiologically isolated syndrome (RIS) with risk factors for conversion to and paraclinical features of MS
  - Spinal cord MRI and the use of Gadolinium are not routinely recommended.

MRI in MS for monitoring disease activity and treatment effectiveness:

- 3D T2w FLAIR and optional Gadolinium enhanced T1w sequences can be sufficient
- The use of Gadolinium is optional and not recommended for all clinical situations. For example, new or enlarging T2 lesions can be used when a recent scan performed within 1 year is available. However, Gadolinium enhanced MRI can be obtained after treatment initiation in the absence of a new baseline scan

1.7 Brain atrophy in MS

According to pathological findings, it has been shown that MS causes significant atrophy in the brain and spinal cord which tends to be more obvious in the progressive types of MS. (31) Researchers have historically used brain and spinal cord atrophy on T1w MRI sequences to investigate the impact of atrophy on disability. Investigators have shown that, regardless of the method used, pwMS have a higher rate of cerebral atrophy when compared to healthy controls (HCs). (69) Even at the earliest stage of MS, people with CIS have shown to develop more brain atrophy in those who develop MS after 1-3 years compared to those who didn't develop MS. (70, 71)

Studies have shown that pwMS who develop higher disease progression tend to also have higher central cerebral atrophy, regional atrophy and whole brain atrophy at baseline, and continue development atrophy over time. (72-74) Grey matter atrophy has also been of interest since the early 2000s, as this has been observed in all stages of MS including CIS, RRMS, SPMS and PPMS. (70, 75-77) Moreover, grey matter atrophy, especially in the cingulate cortex and thalamus, has been shown to be associated with higher neurological impairment scored by both EDSS and MS functional composite, more significantly than white matter atrophy (71, 77)

Brain atrophy measures correlate with neurological impairment and appear to be reproducible, they have therefore been frequently included as endpoint outcomes when assessing therapies. In this thesis we aim to use brain volume measures to assess the impact on prediction of treatment response.

1.8 MS diagnosis

Clinical symptoms, signs and investigations, such as MRI brain/spine scans and lumbar puncture, are required to diagnose MS and exclude other possibilities. The McDonald Criteria was first introduced in 2001 to aid diagnosis and has since been revised 3 times to not only improve MS diagnosis but also make the diagnosis much earlier. (63, 78-81). See Table 3: The McDonald Criteria for MS Diagnosis 2017 for more details. The latest revision applying the McDonald criteria of 2017 have shown higher sensitivity (68-100%) in diagnosing MS after the first clinical attack (clinically isolated syndrome) compared to the 2010 criteria. However, the specificity was low (14-63%) which is expected considering the limitations of the studies including the low number of patients and the short follow-up period. (81-85)

The first neurological episode is referred to as clinically isolated syndrome (CIS) where patient symptoms and clinical findings relate to a focal or multifocal inflammatory demyelinating episode in the central nervous system. This develops acutely or sub-acutely, for at least 24 hours, with or without recovery. This is considered in the absence of fever or infection. CIS can be mono-focal or multi-focal depending on whether the pathology is in one or more locations. Typical CIS presentations include unilateral optic neuritis, focal supratentorial syndrome, focal brainstem or cerebellar syndrome, or partial myelopathy. These presentations are also referred to as relapses, exacerbations or attacks if they occur in patients with a previous history of CIS or those already known to have MS. (16, 80)

MRI brain and spinal cord imaging have played a significant part in the diagnostic criteria as they can establish dissemination in space through multifocal involvement of the central nervous system and/or dissemination in time by showing new and old lesions. Dissemination in space (DIS) can be demonstrated by one or more T2-hyperintense lesions that are characteristic of multiple sclerosis in two or more of four areas of the CNS: periventricular, cortical or juxta-cortical, and infratentorial brain regions, and the spinal cord. Dissemination in time (DIT) can be demonstrated by the simultaneous presence of gadolinium-enhancing and non-enhancing lesions at any time or by a new T2-hyperintense or gadolinium-enhancing lesion on follow-

up MRI, with reference to a baseline scan, irrespective of the timing of the baseline MRI.

There are some limitations to the latest diagnostic criteria, particularly in patients presenting with atypical presentations of MS and non-specific MRI brain lesions that could potentially lead to misdiagnosis. It is crucial that the criteria should only be applied to patients who present with a typical clinically isolated syndrome suggestive of an inflammatory demyelinating episode, and when alternative diagnoses have been excluded. Another limitation to the criteria is the inability to identify the secondary progressive phase of MS objectively, as this is generally diagnosed on the basis of irreversible disability progression that is independent of relapses. One of the first objective definitions was introduced in 2016 defining SPMS as a disability progression by 1 EDSS step in patients with EDSS ≤5.5 or 0.5 EDSS steps in patients with EDSS ≥6 in the absence of a relapse, a minimum EDSS score of 4 and pyramidal functional system score of 2 and confirmed progression over ≥3 months. This definition reached accuracy of 87% compared to a consensus diagnosis by three independent (neurologists) raters. (86)

Table 3: The McDonald Criteria for MS Diagnosis 2017		
Number of clinical episodes	**Number of MRI lesions with objective clinical evidence**	**Additional data required for diagnosis**
CIS		
1	≥ 0	symptoms and examination reflecting a focal or multifocal inflammatory demyelinating episode, developing acutely or sub-acutely, duration of at least 24 h, with or without recovery, and in the absence of fever or infection
RRMS		
≥ 2	≥ 2	Nil
≥ 2	1	DIS shown by another clinical episode at a different CNS location or by MRI
1	≥ 2	DIT shown by another clinical episode, MRI, or unmatched OCBs in CSF.
1	1	DIS shown by another clinical episode at a different CNS location or by MRI and DIT shown by another clinical episode or by MRI or presence of unmatched OCBs in CSF.
PPMS		
Disability progression over 1 year independent of clinical episode. *Plus 2 of the following:* • 1 or more T2-hyperintense lesions typical of MS in 1 or more of the following brain lesions: periventricular, cortical or juxta-cortical, or infratentorial • 2 or more T2-hyperintense lesions in the spinal cord • Presence of unmatched OCBs in CSF		
Abbreviations: CIS = clinical isolated syndrome, DIS = dissemination in space, DIT = dissemination in time, OCBs = oligoclonal bands, CSF = cerebrospinal fluid. PPMS = primary progressive MS		
Adapted with permission from Professor Alan Thompson. (80)		

1.9 MS treatment and response

MS management requires a multidisciplinary team approach. This is because it includes the management of an acute relapse, the use of disease modifying therapies (DMTs) to reduce frequency of or stop relapses, medication for symptom control, psychological support, comorbidity management, and lifestyle adjustments.

Acute moderate to severe relapses are usually managed with high dose oral or intravenous steroids for 3-5 days to speed up recovery. Oral steroids were shown to be non-inferior to intravenous methylprednisolone and had a similar safety profile when used for the management of an acute MS relapse. (87, 88) The current guidelines by the national institute for health and care excellence (NICE) recommends at least daily oral methylprednisolone 500 mg for 5 days. (89) NICE recommends consideration of intravenous methylprednisolone 1000 mg for 3-5 days in patients who previously failed to respond or were intolerant to oral steroids and in those of severe symptoms that require hospitalisation and monitoring. (89) Plasma exchange may be considered for relapses that are severe and refractory to steroids. (90)

PwMS suffer a wide range of symptoms that are important to treat and control for maintaining function and quality of life. The most common symptoms experienced by pwMS and their possible management options are summarised in Table 4: Common symptoms experienced by pwMS and their management. (91)

Prior to the approval of DMTs in MS, several immunosuppressive therapies were frequently given including corticosteroids and steroid-sparing agents: Methotrexate, Azathioprine, Mycophenolate Mofetil, and intravenous immunoglobulin. However, the effectiveness of these immunosuppressants were not studied extensively and did not show significant outcomes. Therefore, they have been used much less often since the approval of DMTs.

Table 4: Common symptoms experienced by pwMS and their management				
Symptoms	**Prevalence**	**Pharmacological management**	**Quality of evidence from clinical trials***	**Non-pharmacological management**
Fatigue	78-94% (92, 93)	Amantadine, Modafinil	Low-moderate (94)	CBT, relaxation therapy, aerobic exercises
Gait impairment	50-91% (93)	Dalfampridine	Low-moderate (95)	Physiotherapy
Urinary dysfunction	90% (96)	Anticholinergics, Botulinum injections	Low (97)	Pelvic floor exercises, intermittent self-catheterisation
Spasticity	84% (98)	Baclofen, Tizanidine, benzodiazepines, botulinum injections	Low-moderate (99)	Exercise, physiotherapy
Bowel dysfunction	52-68% (100)	Stool softeners, stimulants, laxatives	Low (101)	Timed bowel evacuation, dietary fibre, hydration, physical activity
Neuropathic pain	40% (102)	TCAs, SSRIs, Anticonvulsants	Low (103)	Exercise, physiotherapy, psychological therapy, nerve blocks

Abbreviations: pwMS = people with multiple sclerosis, CBT = cognitive behavioural therapy, TCAs = tricyclic antidepressants, SSRIs = selective serotonin reuptake inhibitors,

* Based on the number of clinical trials, the number of participants per trial, and the results obtained in the different trials

Interferon beta-1b was the first DMT approved in 1993. Since then, there has been over a dozen DMTs approved in the UK not only for RRMS but also for PPMS and, more recently, SPMS. The rest of the DMTs that have been developed and used in the last two decades are summarised in Table 5: DMTs; effectiveness, mode of administration, monitoring, and adverse events. (91)

Table 5: DMTs; effectiveness, mode of administration, monitoring, and adverse events						
DMT	**ARR reduction***	**Route and frequency of administration**	**Baseline investigations**	**Monitoring**	**Common adverse events**	**Rare but serious adverse events**
Interferon beta	28-34% (104-107)	S/C: alternate days or x3/week; IM: x1/week or x1/fortnight	FBC and LFTs	FBC and LFTs every 6 months	Headache Flu-like symptoms Leukopenia	Liver toxicity
Glatiramer Acetate	29% (108)	S/C: once daily or x3/week	Nil	Nil	Injection site reaction	Skin necrosis
Teriflunomide	31% (109)	Orally, once daily	FBC, LFTs, blood pressure, QuantiFERON-TB gold, PPD	LFTs monthly for first 6 months, 6 monthly then after.	Headache, deranged LFTs, diarrhoea, nausea, alopecia	Hepatotoxicity, teratogenicity
Dimethyl Fumarate	44% (110)	Orally, twice daily	FBC and LFTs	FBC and LFTs every 6 months	Skin flushing, diarrhoea, nausea, abdominal pain, vomiting	Infections, liver toxicity, lymphopenia, PML
Fingolimod	54% (111)	Orally, once daily	FBC, LFTs, VZV antibodies, fundoscopy, ECG, FDO	FBC and LFTs 6 monthly. Fundoscopy 3-4 months after initiation. Regular skin checks	Headache, deranged LFTs, back pain, Hypertension	Infections, PML, Macular oedema, liver toxicity, hypertension, bradyarrhythmia, heart block, skin cancer
Natalizumab	68% (112)	IV or S/C, 4-6 weekly	FBC, LFTs, JCV serology, MRI brain	JCV serology 3-6 monthly, MRI brain yearly if JCV-negative	Headache, fatigue, arthralgia, UTI, respiratory tract infection	PML, Hepatotoxicity, herpes infections, hypersensitivity reactions
Alemtuzumab	55% vs Interferons (113)	IV, 2 courses 12 months apart	FBC, UEs, TFTs, LFTs, hepatitis panel, VZV antibodies, urinalysis, QuantiFERON-TB gold, PPD	FBC, UEs, and LFTs monthly for 48 months. TFTs every 3 months for 48 months, skin examination yearly	Rash, headache, infusion reaction, thyroid disorder, herpes infection	Autoimmune disorders, HPV infection, stroke, TB, PML
Cladribine	58% (114)	Two treatment courses orally 12 months apart	FBC, LFTs, blood pressure, QuantiFERON-TB gold, PPD, hepatitis panel, age-appropriate cancer screening	FBC and Age-appropriate cancer screening 2 and 6 months after initiating each course	Respiratory tract infection, headache, lymphopenia, nausea	Malignancy, pulmonary tuberculosis, herpes infections, PML
Ocrelizumab	47% vs Interferons (115)	IV, every 6 months	FBC, LFTs, hepatitis panel	QuantiFERON-TB gold, PPD at baseline, FBC and LFTs yearly	Infusion reactions, respiratory tract infection,	Hepatitis B reactivation, PML

					herpes infection	
Ofatumumab	59% vs teriflunomide (116)	S/C, monthly	Hepatitis panel, serum immunoglobulins	Nil	Infections, injection reaction, headache	Hepatitis B reactivation, PML, reduction in immunoglobulins
Ponesimod	30.5% vs teriflunomide (117)	Orally, once daily	FBC, LFTs, VZV antibodies, fundoscopy, ECG	FBC and LFTs 6 monthly. Fundoscopy 3-4 months after initiation. Regular skin checks	Viral infections, deranged LFTs, headache, hypertension	Bradycardia, macular oedema, seizures, skin cancer
Siponimod		Orally, once daily	CYP2C9 genotype, FBC, LFTs, VZV antibodies, fundoscopy, ECG, FDO	FBC and LFTs 6 monthly. Fundoscopy 3-4 months after initiation. Regular skin checks	Headache, hypertension, deranged LFTs	Bradycardia, heart block, liver toxicity, macular oedema, skin cancer

Abbreviations: ARR = Annual relapse rate, S/C = subcutaneous, IM = Intramuscular, IV = intravenously, FBC = full blood count, LFTs = liver function tests, TFTs = thyroid function tests, UEs = urea and electrolytes, ECG = electrocardiogram, FDO = first does observations, PPD = purified protein derivative, VZV = varicella zoster virus, JCV = John Cunningham virus, HPV = human papillomavirus, TB = tuberculosis, UTI = urinary tract infection.
*ARR reduction versus placebo unless otherwise stated.

In general, DMTs primarily aim to minimise MS activity caused by neuroinflammation, hence the majority have been tested in and licensed for RRMS. It is well established that regardless of which therapy, starting any DMT reduces MS activity; clinical relapses and new lesions on MRI, and subsequently reduces long-term disability. (118, 119) Choosing the first DMT can be hard and challenging as the more potent the DMT, the higher the potential risk of severe adverse events. The decision therefore must be based at an individual patient level and taking into consideration several factors including the number of relapses, MRI activity, the efficacy of the DMT, side effects profile, route of administration, and the patient's life circumstances and family planning.

There are two ways of initiating DMTs in MS; the escalation method, where a less effective DMT is started first and then switched to a high-efficacy drug if disease activity persists. This method aims to identify the best control of MS activity with the least risky therapy. The induction method uses high efficacy DMT at onset aiming to control MS activity as early as possible. With the availability of multiple high-efficacy DMTs, routine screening and growing evidence of improved outcomes when used

early are driving the latter method to becoming used more frequently. Trials assessing these therapeutic approaches are in progress, but results will only be seen in years to come. However, retrospective registry studies already suggest that early use of high-efficacy versus lower-efficacy DMT may delay disability progression or transition to SPMS. (119-121) This is potentially explained by the ongoing clinical and subclinical MS activity when using a low efficacy DMT which may lead to irreversible disability and disease progression that could have otherwise been stopped or minimised earlier.

Despite the presence of many therapies, RRMS patients tend to change DMTs within the first couple of years due to poor effectiveness and/or intolerance of side effects. First line injectable and oral therapies, such has interferon beta, glatiramer acetate, dimethyl fumarate and teriflunomide, have a moderate efficacy of 30-40% in reducing relapses, a high rate of immediate adverse events and are less well tolerated. (see Table 5: DMTs; effectiveness, mode of administration, monitoring, and adverse events). In a large retrospective analysis of more than 10,000 RRMS patients in the French nationwide health data system, Vermersch et al found that treatment compliance with dimethyl fumarate, teriflunomide and interferons/glatiramer acetate were 55%, 60% and 39% respectively at 2 years. (122) They also found that 22%, 20% and 25% of patients experienced at least one relapse within the year of commencing dimethyl fumarate, teriflunomide and injectables respectively. In a multicentre retrospective Italian study by Sacca et al. they investigated nearly 3,000 RRMS patients who commenced DMT and showed nearly half of patients switched DMT at least once in 3 years, the majority switching due to lack of effectiveness judged by disease activity, defined as further relapses, new MRI brain MS lesions and/or EDSS progression. (123) Moreover, switching was more frequently observed in those taking lower efficacy injectables compared with those treated with higher efficacy therapies such as fingolimod and natalizumab. Maurer et al. also looked at nearly 600 RRMS patients in Germany across 50 sites commencing first line injectable and oral therapies between 2014-2017 and found that more than 60% of patients had at least one relapse in the first year before

switching to a higher efficacy therapy. (124) More than half of patients switched due to lack of effectiveness and one in five patients switched due to adverse events.

Second line or highly effective therapies (60-80%), including natalizumab, alemtuzumab, cladribine and ocrelizumab, are more effective in reducing MS activity, and although some are associated with infusion related reactions, they are generally well tolerated. However, they carry higher risk of serious adverse events, as summarised in Table 5: DMTs; effectiveness, mode of administration, monitoring, and adverse events. What makes it more difficult is the lack of head to head comparisons of effectiveness of different therapies, particularly in the highly effective group of DMTs. This is why clinicians constantly have to balance the efficacy of treatment versus the risks for each MS patient.

Studies have shown that DMTs also reduce disability progression in MS. (125, 126) In clinical trials, disability progression is commonly defined as an EDSS ≥1.5-points increase in patients with a baseline EDSS 0, EDSS ≥1.0-point increase in patients with baseline EDSS ≥1.0 or ≥0.5-point increase for patients with a baseline EDSS ≥5.5 that persisted for at least 12 weeks. In placebo-controlled trials, first line therapies, interferons for example, have shown a significant reduction of disease progression in RRMS patients at 48 weeks versus placebo (HR 0.62 95%CI 0.40–0.97, p = 0.03). (106) RRMS patients receiving glatiramer acetate, the proportion of patients with improved, unchanged, or worsened EDSS by ≥1.0 at 2 years, were significantly more than those receiving placebo (p = 0.03). (108) In another trial, the proportion of RRMS patients with confirmed disability progression was higher with placebo versus teriflunomide at 2 years (27.3% vs 20.2%, p=0.03). (109) On the other hand, RRMS patients taking dimethyl fumarate did not show a significant reduction in disability progression compared to glatiramer acetate or placebo at 2 years. (110)

Higher-efficacy therapies have also shown reduction in disability progression in RRMS in randomised placebo-controlled trials. RRMS patients taking fingolimod showed significantly reduced risk of disability progression over 2 years versus placebo (HR 0.70; p=0.02). (111) Natalizumab reduced the risk of disability progression in RRMS

patients by 42% over 2 years (HR 0.58; 95%CI 0.43-0.77; p<0.001). (112) RRMS patients taking interferons had significantly higher disability progression compared to those taking alemtuzumab (HR 0.58, 95% CI 0.38-0.87, p=0.008). (127) Cladribine showed significantly lower risk of disability progression in RRMS patients (HR 0.67; 95% CI 0.48-0.93, p=0.02) at 96 weeks compared to placebo. (114) Ocrelizumab showed a significantly lower percentage of RRMS patients with disability progression versus interferons at 96 weeks (9.1% versus 13.6%; HR 0.60; 95%CI 0.45-0.81; p<0.001). (115) Most recently, ofatumumab showed a lower percentage of RRMS patients with disability progression compared to teriflunomide (10.9% versus 15.0%; HR 0.66, p=0.002). (116) Ponesimod on the other hand showed no significant difference in reducing disability progression in RRMS patients at 2 years when compared to teriflunomide. (117)

DMTs have also shown reduction in progression in progressive MS patients, clinically and radiologically. In 2019, ocrelizumab was approved to be the only DMT for active PPMS, as it showed a modest reduction of confirmed disability progression at 12 weeks (32.9% versus 39.3%; HR 0.76, 95%CI 0.59-0.98, p=0.03) and 24 weeks (29.6% versus 35.7%, HR 0.75, 95%CI, 0.58-0.98, p=0.04) compared to placebo. (128) In a sub-group analysis with PPMS patients with active MRI, the treatment effect with confirmed disability progression at 12 weeks reached statistical significance (HR 0.68, 95%CI 0.46-0.99, p=0.04), but did not reach statistical significance at 24 weeks (HR 0.71, 95%CI 0.47-1.06, p=0.09). NICE has therefore recommended the following eligibility criteria for PPMS patient to commence ocrelizumab; aged 18-55 years old, EDSS <7.0, disease duration of less than 15 years, if EDSS >5.0, or less than 10 years if EDSS ≤5.0, with evidence of MRI activity; presence of T1 Gd-enhancing lesions or new T2 lesions. (129) Such restrictive criteria meant that the majority of patients are not eligible and therefore will not benefit.

Most recently siponimod was approved for the management of SPMS. In a double-blind, randomised, phase 3 trial 26% of patients receiving siponimod and 32% of patients receiving placebo had confirmed disability progression (HR 0.79, 95%CI 0.65-0.95; relative risk reduction 21%; p=0.013). (130) Considering the modest

effectiveness, NICE has recommended siponimod for treating SPMS patients with active disease evidenced by relapses or the presence of inflammatory activity on imaging. (131) This again means a large number of patients would not be eligible nor benefit.

1.10 Conclusion

MS is a chronic inflammatory disease of the nervous system associated with physical and cognitive impairment and other symptoms that affect quality of life. The cause of MS remains largely unknown however multiple environmental and genetic factors have found to play a role in developing MS. Inflammatory demyelinating plaques in the brain and spinal cord are the hallmark of pathology in MS, the accumulation of which causes neurological impairment. This is usually assessed using clinical and radiological evaluations which have allowed clinicians and researchers to measure neurological function, disability and the impact on quality of life.

The McDonald Criteria was developed in 2001 to help with the diagnosis of MS and has since been revised 3 times, the last of which occurred in 2017, to improve the sensitivity and specificity of diagnosing MS. In the latest version MS diagnosis can be made earlier and therefore DMTs can be commenced much quicker resulting in a better chance of reducing disease activity and accumulation of disability.

DMTs were first approved in the 1990s and to date there are more than a dozen DMTs approved in the UK for the management of MS. This has given clinicians and patients a varied range of options to consider according to efficacy, risks and life style choices. Together with the earlier diagnosis this has improved the management of MS significantly.

Despite the large number of DMTs currently available we are still unable to predict whether a patient with MS will respond to a particular therapy. Patient and neurologist specific factors of preference (for example, the mode of administration of the drug), disease type and activity and access to medications play a role in the decision-making process of choosing therapy for the individual MS patient. However, there is currently no tool to guide patients and clinicians in selecting the optimum DMT for a particular patient. To date it has been challenging to predict which DMT will be most effective whilst minimising side effects due to the complex number of factors needed to be considered. In recent years however, machine learning has been used in studying vast quantities of complex information, developing algorithms to

find patterns and identifying answers to complex questions. In the next chapters of this thesis the emphasis is going to be on developing a better understanding of predictors in MS diagnosis and management, including the use of machine learning, and working towards personalised medicine and customising therapies for the individuals with MS.

Chapter 2. Machine learning in MS

2.1 Introduction

Standard care of MS patients generates a wealth of clinical data including clinical symptoms and signs, imaging data, and laboratory data. Due to the heterogeneity of onset, diagnosis, classification, treatment response and prognosis, it is difficult to predict outcomes in those with MS. Historically an older age at the onset of MS, as well as being of the male sex, have been associated with poor prognosis. (132) Early clinical and radiological predictors associated with poor prognosis include spinal cord presentation, higher EDSS at baseline, a greater number and larger volume of T2w lesions, and brain atrophy. (133) Other factors which may have value in clinical practice to predict the risk of MS activity and disability are summarised in **Error! Reference source not found.**. These include demographics, clinical characteristics, MRI measures and laboratory investigations.

This extensive amount of work has been invaluable for understanding the course of the disease and has led to the development of a continually changing treatment landscape over the last three decades. Despite this, disease mechanisms and the RRMS course continue to be heterogeneous and unpredictable. This unpredictability has provoked the introduction of artificial intelligence and machine learning in the field of MS over the past two decades. Therefore, the focus of this chapter will be to review the literature on the use of machine learning in MS. However, firstly we will discuss the basics of machine learning, types and methods of evaluation, and then we will review the literature.

Table 6: Features associated with poorer clinical outcomes in MS		
	Features	**Outcome measures**
Demographics	Age at MS onset	Disability as per EDSS: Those aged 20-29 years at MS onset reached EDSS 6.0 earlier compared to those aged 40-49; HR 0.55 (0.42-0.73) vs 0.13 (0.10-0.19), 95%CI, p<0.001 (134)
	Sex	The median time from MS onset to reaching disabling EDSS was significantly longer in females; e.g. reaching EDSS 6.0: 19.2 (11.6-31.7) years in males vs 24.3 (14.5-36.0) years in females, p-value <0.0001. (135)
	Ethnicity	MENA-MS descent showed greater disability progression versus EUR-MS over time (EDSS change = 0.24 vs. 0.06, p = 0.01) and a higher MSSS (3.12 vs. 2.67, p = 0.04). (136)
	Cardiovascular co-morbidities	The median time from MS diagnosis and needing ambulatory assistance was 18.8 years in patients without and 12.8 years in patients with vascular co-morbidities. (137)
	Psychiatric co-morbidities	The presence of psychiatric co-morbidities was associated with a higher EDSS score (β coefficient = 0.28, p = 0.0002, adjusted for disease duration and course, age, sex, socioeconomic status, physical comorbidity count, and DMT exposure) (138)
Clinical	Number of relapses in the early years of MS onset	≥2 relapses in the first year of MS was significantly associated with a shorter time until reaching EDSS 6.0; HR 2.8 (1.60-4.93), 95%CI, p<0.001. (139)
	Frequent attacks with slow recovery	Attack frequencies, in the first 2 years, of 1 versus >or=3, gave differences of 7.6, 12.8 and 20.3 years in times from MS onset to EDSS 6.0, 8.0 and 10, respectively. (140)
	Spinal cord presentation at onset	Spinal cord presentation at MS onset was predictive of higher risk progression to SPMS (adjusted OR=2.01; p=0.06) (141)
MRI	Gadolinium-enhancing, infratentorial and spinal cord lesions	Baseline gadolinium-enhancing (OR 3.16, p<0.01) and spinal cord lesions (OR 4.71, P < 0.01) in CIS are independently associated with SPMS at 15 years. (7) ≥2 infratentorial lesions predict long-term disability (HR 6.3), reaching EDSS ≥ 3.0. (142) Spinal cord lesions are associated with more relapses and worse outcomes, EDSS, at mean follow up 37.2 months (143)
Laboratory test	Presence of unmatched OCBs at presentation	Increased the risk of CDMS [adjusted HR 1.3 (1.0-1.8)] and disability [adjusted HR 2.0 (1.2-3.6)], reaching EDSS ≥ 3.0. (144)
Abbreviations: MS = multiple sclerosis, EDSS = expanded disability status scale, HR = hazard ration, MSSS = MS Severity Score, MENA-MS = MS patients of Middle Eastern and North African descent, EUR-MS = MS patients of European descent, DMT = disease modifying therapy, OR=Odds Ratio, SPMS = secondary progressive MS, CIS = clinically isolated syndrome, OCBs = oligoclonal bands, CDMS = clinically definite MS.		

2.1.1 Artificial intelligence and machine learning

Artificial intelligence (AI) and machine learning (ML) algorithms are able to detect and compactly represent important patterns within large and complex clinical data. This ability can be used to develop data-driven tools to stratify patients with possible consequences for their care including diagnosis, management, monitoring and response to therapy. (145)

AI and ML are loosely defined. Here, we consider AI to be a set of methods that can emulate smart human behaviour, such as making decisions under uncertain conditions. ML, which is a branch of the wider field of AI, studies algorithms that enable performing special tasks and inferring patterns from data. (146) It is particularly useful when working with very large datasets such as MRI scan images. It works well in performing tasks that require repetitive routine activity (such as interpreting MRI images) and, depending on the task, can possibly complete tasks faster and more accurately than a human interpreter. However, ML relies on sufficiently abundant data to ensure the model is robust and generalisable to other unseen data.

2.1.2 Machine learning types

There are many types of ML, two of the most common types are supervised and unsupervised learning, demonstrated in Figure 3: Machine learning algorithms.

In supervised ML an algorithm is trained on a training cohort to learn patterns associated with the specific category, such as healthy control or treated cohorts. Once patterns are learnt from the data given, ML algorithms assign categories to an unseen test cohort. Supervised ML algorithms are commonly used for two types of questions; classification and regression. Classification is used to predict which group an observation belongs to, for example patients vs healthy controls. Regression is used to predict a continuous measure, for example time to diagnosis/disability. Examples of supervised ML techniques include: neural networks, decision trees,

random forest and support vector machine. This is summarised in Table 7: Common Supervised ML techniques (147)

In unsupervised ML, training data are not categorised or labelled. Instead the algorithm attempts to find patterns in the data by identifying clusters created from the similarity in the data. Examples of unsupervised ML methods include hierarchical clustering.

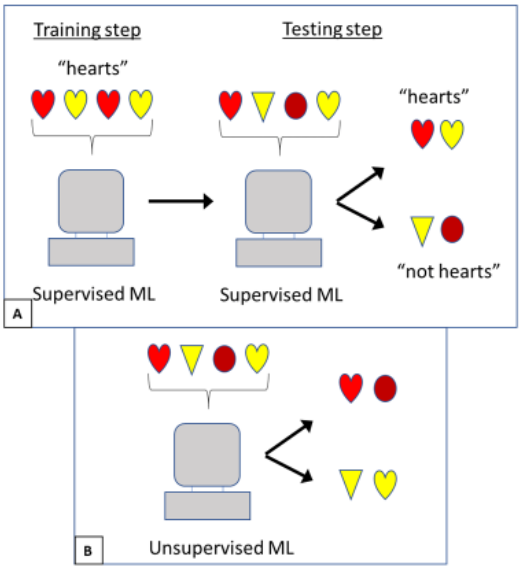


Figure 3: Machine learning algorithms.
A. Supervised machine learning (ML) in which the training step includes training data with labels to create a predictive model. The testing step includes showing unlabelled data to the predictive model to identify, in this example, hearts from other shapes.
B. Unsupervised ML in which unlabelled data are processed by the model to see what patterns can be shown, grouping red shapes in this example.

Table 7: Common Supervised ML techniques	
Random forest	Grows a large number of decision trees, each sees only a subset of data, and learns output from input by combining the predictions
Decision tree	Learns outputs from inputs given by a series of yes/no questions in a tree-like structure where internal nodes represent the variables of a dataset, branches represent the decision rules and each leaf node represents the outcome.
Support vector machine	A binary classification technique that can be modified to multiclass classification or regression. They look to split the predictor space into two, so that data points from each class are concentrated on one side of the decision boundary
Neural networks	learns outputs from inputs through a series of nested nonlinear function, encoded in a network of "neurons", which may vary in its topology.

2.1.3 Artificial neural networks

Artificial neural networks (ANNs) are biologically-inspired programming paradigm, resembling the networked neurons in the human brain. They are built by three types of layers: An input layer, whose neurons connect into any number of hidden layers, which in turn feed forward to an output layer, as demonstrated in **Figure 4: A simple artificial neural network**. Neurons have multiple connections with neurons in the other layers with different weights depending on their connections. The neurons and layers have a fixed structure, but the networks learn by adjusting the weights depending on the outcomes. Backpropagation is a method commonly used for adjusting weights depending on the amount of error between the ANN outcome and the actual outcome, termed prediction error. Strengthening or weakening the weights depends on whether the true outcome was unexpectedly positive or negative. Therefore, the size of prediction error guides how weights are changed. (148)

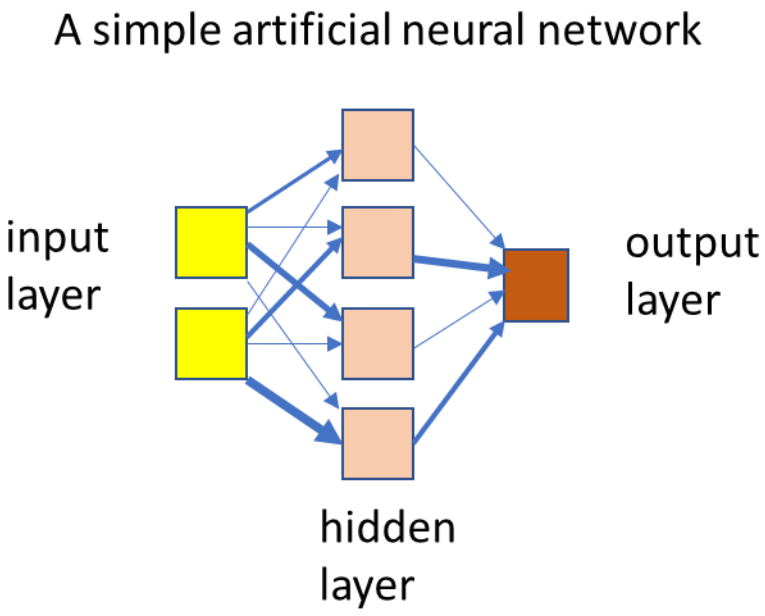


Figure 4: A simple artificial neural network
Artificial neural networks (ANNs) contain neurons (squares) that are organised into an input (yellow), output (brown) and any number of hidden (pin) layers, one layer in this example. The blue lines represent the different weights of connecting different neurones, demonstrated by the thickness of the lines.

Slight variations in the structure of an ANN can help in performing different tasks. Adding connections, to and from the same neurons, within a hidden layer create recurrent neural networks. These address sequences of information to help in tasks such as speech analysis. Convolutional neural networks (CNNs) are used to capture relevant features from an input, as they filter and simplify complex inputs into the most important features that can allow for more efficient data processing. Another form of ANNs includes deep learning algorithms that have many hidden layers, which allow them to perform more complex operations.

2.1.4 Machine learning model evaluation

ML models are situation specific and dependent on size, type, and dimensionality of data. There are many ways of evaluating models' performances, as summarised in Table 8: ML model evaluation. (147) ML models are complicated and require optimisation of parameters to capture the best relationship between parameters and outcome. Capturing many errors from the training data points and failing to find the underlying trend of the data cause poor performances, known as "underfitting". On the other hand, "overfitting" is an example of a model that captures no errors in the training data because the model is built on each deviation in the data point and becomes too sensitive as it captures patterns that only present in the training dataset. Such model would not perform well in testing external datasets, i.e. not generalisable. It is therefore essential for any ML method to account for this by training the model on a large training dataset and test the model generalisability in an external independent dataset. This process is often difficult as it requires extensive collaboration, standardisation of variables and approval for data sharing across different institutions and countries. This is why most studies use cross validation.

Cross validation is a technique that involves splitting the data into a training dataset and a testing dataset. (149) A training dataset is used to train and optimise the model and subsequently the model performance is evaluated on the unseen testing dataset. There are many methods used to perform cross validation including leave-one-out-cross-validation (LOOCV) which reserves one data point from the dataset and train the model on the remaining data. This process is repeated for each data point.

Although this method helps is making use of most data points and minimise bias, it can take a long time to process if there are many data points and can give a high variation in testing model effectiveness, as the testing is performed on and influenced by one data point only. K-fold cross validation is another method that works by splitting the dataset into k"folds". (150) For each k-fold in the dataset, the model is built on k-1 folds of the dataset. The mode is then tested to check the effectiveness for kth fold. The error on each prediction is recorded during this process. This process is repeated until each of the k-folds is used as the test dataset. The average of the k recorded errors is known as the cross-validation error, which is used as a performance value for the model. It is worth noting that a lower number of k can lead to more bias and therefore not desirable, while higher number of k means less bias, the model can suffer from large variability.

Table 8: ML model evaluation	
Accuracy	is the percentage of correct predictions
Balanced Accuracy	is a measure of the total number of correct predictions in either class, taking into account an unbalanced dataset
Area under the receiver-operator curve (AUC)	is used for binary classification problems, using a plot of sensitivity versus specificity to determine model performance at various threshold settings.
Precision	is equivalent to positive predictive value
Recall or Sensitivity	identifies true positives correctly
F-score	is an accuracy measure calculated using precision and recall
R^2	Is a statistical measure that represent the proportion of variance in the dependent variable that is explained by the independent variable(s) in the regression model.

2.1.5 Frequentist and Bayesian statistics

Complex computational models can have different forms and focus and can be categorised along different axes. Some are principled and based on probability theory (Probabilistic) or even full probability distributions (Bayesian). Others focus on algorithms that try to find the best answer in a possible search space without regard for uncertainty. Machine learning algorithms typically fall int the latter category, but some types of machine learning can be Bayesian. The same regression model can be estimated using a standard (frequentist) statistical algorithm (e.g ordinary least squares), a more complex non-linear machine learning algorithm (such as a support vector machine) or a fully probabilistic Bayesian algorithm (such as Markov chain Monte Carlo). It is therefore important to understand the differences and when to choose different methods – here we focus on Bayesian methods. This requires a brief review of the Probability theory which lies at the foundation of many methods.

Probability measures the uncertainty of outcomes of a random variable. In most studies there are many random variables used that can interact in complex ways. The probability of two variables can be mathematically related in a number of ways using joint, marginal and conditional probability rules. The probability of two events occurring at the same time (even A=a and event B=b) is their joint probability. Marginal probability is the probability of an event irrespective of the outcome of another variable (event A=a, averaged over all variations of B=b). The probability of one event occurring in the presence of a second event (event A=a given we know that event B=b) is conditional probability. (151) Complex predictive probabilistic models are derived from these principles.

Frequentist statistics estimates the probability of an event occurring when the experiment is repeated in controlled conditions to assess the outcome. It is often used to test if the null hypothesis can be rejected or not. In clinical trials, the most common frequentist decision procedure uses p-values, typically needed to be lower than a pre-set arbitrary cut-off (e.g. alpha = 0.05) given a particular sample and design. One of the common problems with the traditional frequentist approach is the statistical decision is tied to the final sample size. More complex statistical

procedures are required if the sample size changes after the start of the experiment. Similarly, if the results were positive, it is not possible to know if the results could have been reached with a smaller sample size (post hoc power calculation) but this data can be used to estimate the power of future trials. (152)

Bayesian statistics is a mathematical method that applies probabilities to statistical questions. (153) It provides the tools to update beliefs in the presence of new evidence or data. It is important to understand conditional probability, which is the probability of an event, A, given another event, B, equals the probability of B and A happening together (joint) divided by the marginal probability of B; (154)

$$P(A|B) = P(A \cap B)/P(B)$$

The joint probability $P(A \cap B)$ in this equation can be re-written as another conditional probability using a re-arrangement of the same equation;

$$P(A|B) = [P(B|A) \text{ X } P(A)]/P(B)$$

The marginal probability of B can be calculated as:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

Since $P(B \cap A_i) = P(B|A_i) \text{ X } P(A_i)$

The equation of conditional probability is therefore;

$$P(A_i|B) = (P(B|A_i) \text{ X } P(A_i))/(\sum_{i=1}^{n}(P(B|A_i) \text{ X } P(A_i)))$$

The Bayesian framework contains the following steps: using available knowledge for a given parameter in a statistical model, before data collection, through the prior distribution; using the information about the parameters in the observed data to

determine the likelihood function; and using the combination of the prior distribution and the likelihood function to determine the posterior distribution. (154)

Since P(A|B) is the posterior probability, P(A) is the prior probability, P(B|A) is the likelihood and P(B) is the evidence, The Bayes Theorem equation is therefore;

Posterior probability = Likelihood X Prior probability / Evidence

This helps in updating one's knowledge by combining prior knowledge and observed data to reach inferences. This is one of the main differences compared to frequentist statistics, using p-values and confidence intervals, as frequentist inferences do not quantify what is known about the parameters.

The need for prior distribution is the focus of the argument against the Bayesian approach. This argument becomes not significant when the sample size increases. However, it is important to understand that choosing the best available data leads to optimistic priors that can influence outcomes.

Both frequentist statistics and Bayesian statistics will be implemented later in this thesis, chapter three and four, respectively. However, having discussed the use of model building using machine learning, the literature review in the next section will demonstrate the use of machine learning methods in the field of MS that would be of interest to neurologists.

2.2 Methods

As part of this detailed, non-systematic, literature review PubMed was used to select the appropriate studies of interest with the following terms: “multiple sclerosis” and “artificial intelligence” or “machine learning” in the title or abstract. Up to April 2021 a total of 7204 studies were initially identified. After exclusions 702 studies were reviewed and only articles deemed relevant to MS neurologists were included. Studies not written in English, or those that had a technical overview and were not applied to MS, were excluded. Included articles were divided and summarised, as per Figure 5: Literature search summary flowchart, in the following sub-sections.

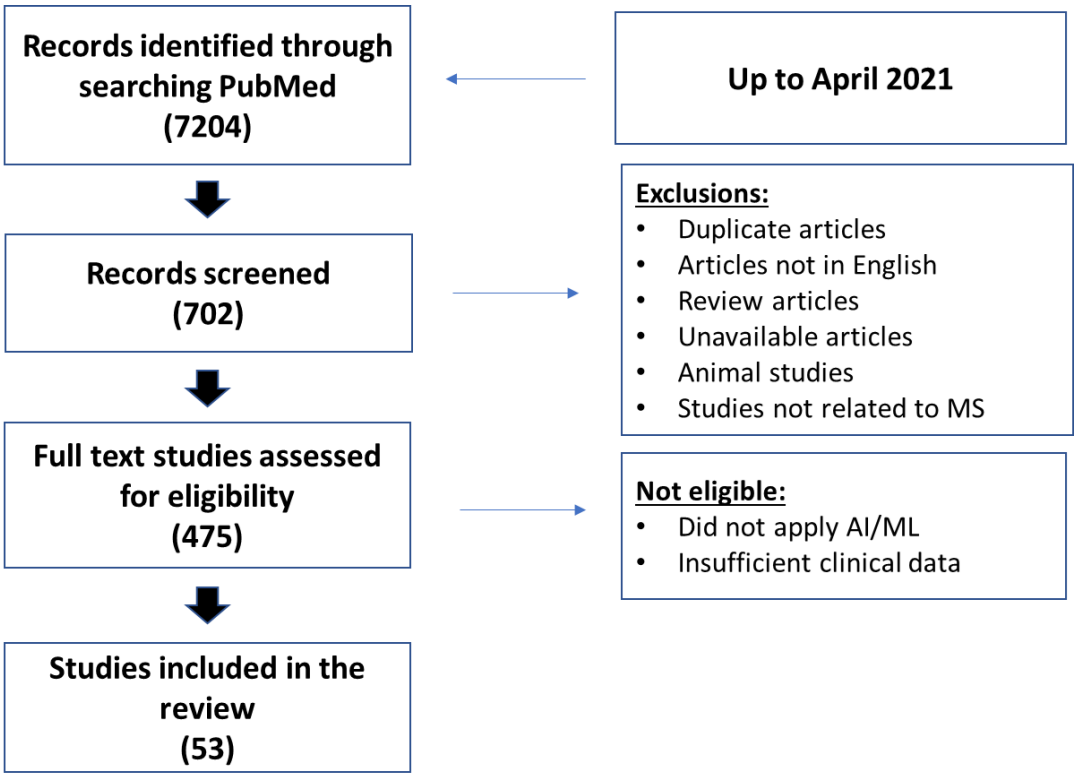


Figure 5: Literature search summary flowchart

AI = artificial intelligence, ML = machine learning, insufficient clinical data = data provided was not clinically relevant, e.g. using AI/ML on health records and insurance claims.

2.3 Results

2.3.1 ML application in MS diagnosis

2.3.1.1 Using MRI

Patients with MS are difficult to assess due to the heterogeneous clinical presentation, symptoms and progression over time. Combining copious amounts of data and ML opens up great opportunities for understanding and monitoring MS using MRI. (155) ML algorithms using MRI metrics have been used in the field of MS diagnosis over the last decade and more extensively in the last 5 years. They aimed to classify MS patients versus healthy controls (HCs), although in most cases they included patients with already established MS diagnosis. The studies that focussed on the use of MRI are summarised in Table 9: ML application in MS diagnosis using MRI.

Initially Weygandt et al. trained a Support Vector Machine (SVM) to predict diagnostic accuracy of 41 RRMS patients and 26 age and gender matched HCs using their MRI (T2w) images. (156) They found that among regions containing lesions a posterior parietal white matter area was more informative with 96% accuracy, after LOO-CV. Among the normal appearing grey matter, the cerebellar regions were more informative with an accuracy of 84%. Finally, among the normal appearing white matter areas, a posterior brain area was most informative with an accuracy of 91%. This study identified vital areas of MS associated tissue changes using ML technique utilising standard MRI sequences. The main limitation of this study is the small sample size and the lack of an external validation cohort.

In 2012 Richiardi et al. used resting state functional MRI (rsfMRI) to develop a predictive model of connectivity alterations in MS. (157) They included 22 MS patients and 14 HCs to build a whole brain connectivity matrix from slow oscillations of region-averaged time series and a discriminant function to indicate which connections are most affected by MS. Strict cross-validation was used as a classification performance which provided a sensitivity of 82% and specificity of 86% to differentiate between MS and HCs. Subcortical and temporal regions, and

contralateral compared to ipsilateral connections, were found to show the most discriminative connectivity. Although the findings suggest that predictive models using rsfMRI can identify changes caused by MS, due to the small sample size this requires further testing before any consideration of potential new markers are utilised clinically. The main limitations for this study are the small sample size and the use of LOO-CV, which gives limited generalisability to an external cohort. This process is also difficult to apply since rsfMRI is not currently used in routine clinical practice.

In 2016 Eshaghi et al. used random-forest classification to differentiate MS from neuromyelitis optica (NMO) and HCs using brain grey matter imaging. (158) They included 90 patients from Iran (25 MS patients, 30 NMO patients, and 35 HCs) and 54 patients from Italy (24 MS patients, 20 NMO patients, and 10 HCs). The volume, thickness and surface of 50 cortical grey matter regions were extracted from T1 and T2/FLAIR MRIs alongside volumes of the deep grey matter nuclei to generate 3 random forest models to differentiate MS, NMO and HC groups. For validation they randomly assigned participants from both centres to either the training or test set, so that each set contained half of the participants. Initially they performed the training step and then the cross-validation on the remaining half (with 5,000 repetitions). The accuracy was compared to the final clinical diagnosis. The model was able to differentiate MS patients from NMO with 74% accuracy. MS patients showed greater atrophy in deep grey matter. Using thalamic volume and white matter lesion volume, the accuracy of the model improved to 80% in distinguishing MS from NMO patients. The model achieved even higher accuracy when classifying MS versus HCs (92%) and NMO versus HCs (88%). This study highlighted that grey matter can be used as a possible biomarker to diagnose MS and differentiate it from NMO using ML and clinical MRI scans. It is important to note that the NMO group were significantly older than the MS group which may have impact on grey and white matter volumes. Although they used participants from two different centres with a more robust validation step, the final model was not tested against an unseen external cohort.

In 2018 Yoo et al. looked at combining myelin and T1-weighted (T1w) MRI features in the NAWM to differentiate pwMS from healthy subjects. (159) Using unsupervised deep learning techniques they extracted myelin maps from 3D image patches and corresponding T1w MRIs and developed a joint myelin-T1w feature representation. The resulting features were used to train a random forest classifier using 55 RRMS patients and 44 healthy subjects in an 11-fold cross-validation test. This method showed an accuracy of nearly 88% in distinguishing MS from HC, which is a higher accuracy compared to assessing T1w (70%) or myelin (84%) images alone. This highlighted a possibility in identifying MRI features in the NAWM that are more sensitive and specific to MS pathology. The main limitations of this study include the small sample size and the application of this model on patients with an already confirmed MS diagnosis. Since the model is based on white and grey matter changes, other neuro-inflammatory conditions may not be differentiated. Finally, the utility of this model is limited due to the lack of testing on an unseen cohort.

Zurita et al. used SVM classifications as a ML technique to characterise RRMS patients from healthy controls using diffusion and functional MRI scans in 2018. (160) They used a novel method to find relevant brain areas and connectivity measures to identify pwMS by extracting fractional anisotropy maps, structural and functional connectivity from diffusion tensor imaging (DTI) and resting state functional MRI (rsfMRI) scans. These in turn created SVM classifiers from 104 RRMS patient and 46 healthy controls. These classifiers were able to differentiate MS from healthy controls with an accuracy of 89%. Importantly, the brain areas that were most helpful to the classification were medial frontal cortices, lingual gyrus, right occipital and left frontal orbital. This method therefore provided better understanding of MS through identifying relevant brain regions with structural and functional connections. One of the main limitations of this study was including RRMS with low disability, median EDSS 1.5, and therefore the model couldn't distinguish between those with higher EDSS. It is also important to note that DTI and rsfMRI scans are not routinely undertaken in clinical practice and it is therefore difficult to assess this model in an external unseen group.

In 2019 Sacca et al. evaluated ML techniques for diagnosing MS using a rsfMRI connectivity data. (161) SVM, Random Forest, Artificial Neural Network, K-nearest-neighbour and Naïve-Bayes were the evaluated algorithms. The independent Component Analysis was used to look at the rsfMRI and select the relevant brain networks evaluated from 18 untreated RRMS patients and 19 matched for demographic variables healthy controls. Fifteen networks were detected and the mean signals were used for the classification. Feature selection tasks were used to choose the relevant variables from all the algorithms and identify the best discriminant network between MS and HCs. SVM and Random Forest were found to be the best approaches with similar accuracies of nearly 86%. This was a positive step towards the application of ML techniques using rsfMRI for clinical diagnosis of MS. The main limitations of this study include the small sample size and the use of rsfMRI, which is not available in routine clinical practice, and makes this method difficult to evaluate in an external independent cohort.

Azarmi et al. also used task-related fMRI modality to investigate the brain's neural substrate and compare them between MS and HCs using ML in 2019. (162). Blood-oxygen level dependent (BOLD) fMRI data during paced auditory serial addition test (PASAT) were collected for 20 participants (8 RRMS patients and 12 HCs). Granger causality analysis (GCA) was used in brain regions' time series on all participants to create a brain network. A combination of Wilcoxon rank-sum test and Fisher score were used to classify MS from HCs using SVM (and LOOCV method) with an accuracy reaching 95%. This was the first study to show a significant difference in the global flow coefficient between MS patients and HCs in several brain regions including Hippocampus, Para Hippocampal, Thalamus, Superior temporal gyrus, Caudate, Medial Frontal Superior Gyrus, Fusiform, Pallidum, and many parts of Cerebellum. This experiment showed a promising ML classification technique in identifying MS patients using task-related fMRI. The main limitation of the study is the small sample size. It is also difficult to reproduce or implement considering fMRI and PASAT are not routinely performed in clinical practice.

Neeb et al. investigated the use of different multivariate supervised ML algorithms for MS diagnosis using quantitative MRI in 2019. (163) They included 52 MS patients and 45 HCs where all participants were scanned using a 3T scanner and the average of T1, T2*, total water content and myelin bound water content were evaluated in grey and white matter. T1 in grey matter and myelin water content in white matter were the best differentiating variables in the multivariate analysis, with cross validated multivariate model reaching accuracy of nearly 84%. However, the performance dropped to around 74% accuracy when artefacts were included. This demonstrated that quantitative MRI sequences can help in automated prediction of MS diagnosis. The limitations of this study included the use of a small sample size and the lack of external validation. Importantly, one of the potential confounding factors was the significant difference in male to female ratio between the MS and HC groups despite being age-matched.

Table 9: ML application in MS diagnosis using MRI					
	Reference	**Population**	**Markers**	**Methods**	**Accuracy**
1	**Weygandt et al. 2011**	42 RRMS and 26 HCs	Standard MRI brain (T2w)	SVM	Posterior parietal (96%), Cerebellar areas among normal appearing grey matter (84%)
2	**Richiardi et al. 2012**	22 MS and 14 HCs	rsfMRI – functional connections	Discriminative connectivity	Subcortical and temporal regions showed highest discriminative connectivity changes, 86%
3	**Eshaghi et al. 2016**	From Iran: 25 MS, 30 NMO and 35 HCs. From Italy: 24 MS, 20 NMO, 10 HCs	MRI brain: 50 cortical grey matter regions and volumes of the deep grey matter nuclei extracted	Random forest	MS Vs HC 92% MS Vs NMO – 80% using thalamic and white matter lesion volumes
4	**Yoo et al. 2018**	55 RRMS and 44 HCs	MRI brain - combining myelin and T1w MRI features in NAWM	Random forest	88%
5	**Zurita et al. 2018**	104 RRMS and 46 HCs	Diffusion tensor imaging (DTI) and rsfMRI	SVM	89%. Medial frontal cortices, lingual gyrus, right occipital and left frontal orbital – most helpful
6	**Sacca et al. 2019**	18 RRMS and 19 HCs	rsfMRI	SVM, Random forest, Neural network, K-nearest-neighbour and Naïve-Bayes	SVM & random forest – 86%
7	**Azarmi et al. 2019**	8 RRMS and 12 HCs	task-related fMRI	SVM	95%
8	**Neeb et al. 2019**	52 MS and 45 HCs	Quantitative MRI; T1, T2*, myeling bound water content in grey and white matter.	Multivariate supervised analysis	T1 in grey and myelin water content in white matter were the best differentiating variables – 84%

Abbreviations: ML = machine learning, MS = multiple sclerosis, MRI = magnetic resonance imaging, RRMS = relapsing remitting MS, HC = health control, T2w = T2 weighted, SVM = support vector machine, rsfMRI = resting state functional MRI, NMO = neuromyelitis optica, NAWM = normal appearing white matter.

2.3.1.2 Using optical coherence tomography (OCT)

Retinal OCT is a non-invasive, rapid, objective and reproducible method used for measuring the retinal CNS layers, which appear to be affected in CNS diseases. Pérez del Palomar et al. used ML methods to detect damage in the retinal nerve fibre layer (RNFL) and the complex Ganglion cell Layer-Inner plexiform layer (GCL+) in people with MS in 2019. (164) The eyes of 80 age-matched RRMS patients and 180 HCs were assessed and decision trees, multilayer perceptron and SVM were the ML techniques used for the analysis of the RNFL and GCL+ thickness loss between the two groups. Decision trees provided the best prediction of MS (97%) using RNFL data in the macular area with 10-fold cross-validation. When a wider protocol was used in the extended area between the macula and papilla the accuracy reached 95%. Using SVM the accuracy of differentiating the two groups was only 91%. This study provided evidence that RNFL thickness measurements, particularly at the macula, can differentiate between MS and HCs using ML. One of the main limitations of this study was the inclusion of only Caucasian participants. Another limitation being the average disease duration of 7 years at assessment and the fact it is important to include those with early disease onset if OCT is to be used as part of the diagnostic tools. An EDSS score was not included in the analysis which could have a relation between MS disability and RNFL or GCL+ thickness.

Cavaliere et al. also looked at the use of OCT in diagnosing MS through ML. (165) They performed OCT on 48 MS patients without a previous history of optic neuritis, and 48 HCs, and the mean of the macular thickness and peripheral areas were calculated and compared between the two groups. A confusion matrix was obtained by SVM as an automatic classifier. The total Ganglion cell layer thickness was the most dominant variable, measured using the peripapillary area and macular retina thickness in the nasal quadrant of the outer and inner rings. After applying LOOCV, the results showed an accuracy of 91%. This study highlighted the possibility of identifying MS patients using OCT with ML techniques that can potentially learn the structural neurodegeneration in the retina. One limitation of this study is the inclusion of only 20% of treatment-naïve patients, which could have a confounding effect on the level

on neurodegeneration in different parts of the retina. Another limitation is that the possible correlation between OCTs findings and MRI results were not studied. Finally, this was a single centre study and was not validated in an external unseen cohort.

2.3.1.3 Using metabolic biomarkers

In 2017 Lötsch et al. used unsupervised ML to analyse lipid markers, pursuing the theory of the most relevant markers that are associated with the MS diagnosis. (166) Three main classes of lipid markers were identified including ceramides, eicosanoids, and lyosophosphatidic acids. ML algorithms identified distinct ceramides serum concentration data structures that separate MS from healthy controls with an accuracy of 94%. This preliminary study provided a possible biomarker in MS. A couple of years later, Lötsch et al. used further unsupervised ML techniques including self-organising maps of neuronal networks to identify 43 serum concentrations of various lipid-markers from 102 MS patients and 301 HCs which provided a basis for a classifier (biomarker). (167) This led to the development of supervised ML algorithms using random forests and computed ABC analysis-based feature selection. Bayesian statistics, used in creating a marker to map diagnostic groups of MS versus HCs, identified a complex classifier comprising of eight serum lipid-biomarkers differentiating pwMS from HCs with an accuracy of 95% in training and test data. This was further supportive to the development of MS serum lipid biomarkers. It is important to note that in this study the use of contraceptives was not considered, which could be a confounding factor to lipid patterns. There was also a significant age difference in the MS group versus the HCs. Finally, the results also needed to be validated in an unseen dataset.

Andersen et al. used blood metabolites to develop biomarkers in diagnosing MS in 2019. (168) Age and body mass index matched 12 white, male, treatment-naïve MS patients and 13 white, male, healthy controls were included. Random forests as a supervised ML technique was used to evaluate 325 metabolites that are important for MS classification. Using logistic regression, they found 12 metabolites to be associated with the diagnosis of MS, of which only 6 achieved an accuracy (AUC) of 80% or more including pyroglutamate, laurate, acylcarnitine C14:1, N-

methylmaleimide, and 2 phosphatidylcholines. Using multivariable regression models, the associations of the top metabolites and the whole-genome expression data were investigated and interestingly, HLA-DRB1*15:01, which is the predominant MS genetic risk allele, was associated with one of the six important metabolites. This study identified important metabolic changes in serum associated with MS, having included genomic and genetic data. This was one of the first studies looking at possibly developing serum metabolite biomarkers for the diagnosis of MS using ML. The main limitations of this study include the small sample size and although the cohort was homogenous for sex, BMI, disease duration and smoking status, there could be associations of these variables that were overlooked. The findings were also not tested on an external cohort.

2.3.1.4 Using immunological markers

In 2017 Ostmeyer et al. applied ML to B lymphocytes repertoires collected from CSF to help in differentiating RRMS from other immune-related neurological diseases (OND). (169) Applying the maximum likelihood optimisation techniques, B cell receptor heavy chain genes in each patient repertoire was used as a classifier to detect RRMS. Using LOOCV, the classifier had an accuracy of 87% in the training cohort (23 participants; 11 with RRMS and 12 with OND). However, using an unseen cohort (102 participants; 60 with RRMS and 42 with OND) from a separate study, the classifier accuracy was 72%. This was a novel technology which paved the way for statistical classifiers to map immune repertoire sequence data to a single diagnosis with a high degree of diagnostic capability. The main limitation of this study is the sample size and that when tested on a larger group of patients the accuracy falls. Not including healthy controls in this study can be an important limitation considering these findings may not be confined to neurological disease. Finally, due to the increasing sensitivity and specificity of clinical findings and radiological investigations, the use of lumbar puncture as an invasive procedure in the diagnosis of MS is no longer commonly used.

In 2019 Goyal et al. used ML methods to differentiate pwMS versus HCs based on serum cytokines. (170) Serum levels of 8 cytokines (IL-1β, IL-2, IL-4, IL-8, IL-10, IL-13,

IFN-γ, and TNF-α) were analysed to predict the diagnosis of MS from two different studies of USA (859 MS patients and 128 HCs) and Russian (92 MS patients and 71 HCs) populations. Further models were developed to classify RRMS from non-RRMS with the addition of further variables including age, gender, disease duration, EDSS and Multiple Sclerosis Severity Score (MSSS). After splitting the data into training (70%) and testing (30%), random forest was the best model (Out of the 4 different models used (SVM, decision tree, random forest, and neural networks)) for MS diagnosis with accuracy reaching 91% after performing 6-fold cross validation. However, the accuracy in differentiating RRMS from non-RRMS only reached 70%. This study highlighted the potential application of ML using serum cytokines levels as predictors of MS diagnosis. Although using different populations, splitting the data into training and testing data and performing cross validation are ways to reduce overfitting. The study remains suboptimal as the ideal test would be to apply this method on an external independent cohort.

2.3.1.5 Using genetics

In 2020 Ghafouri-Fard et al. used artificial neural network (ANN) to predict the risk of MS based on single nucleotide polymorphism (SNP) genotypes. (171) They evaluated genetic data of 401 MS patients and 390 age-/sex-matched HCs where 23 SNPs within 11 genes were assessed in all participants. The DD genotype of the *ACE* rs1799752 had the most prominent risk of MS, while the TT genotype of the *GAS5* rs2067079 was the most protective against MS. This ANN based model provided an accuracy of 65% following 5-fold cross-validation, which is suboptimal, however this is a preliminary study creating ANN on genomic data to predict risk of MS using SNP genotypes. Although the diagnostic power may be improved with incorporation of further genomic data, it is important to test this model in an external unseen cohort.

2.3.1.6 Using gut microbiota

In 2018 Forbes et al. conducted a pilot experiment to explore the association of the gut microbiota with immune mediated inflammatory diseases (IMIDs) such as MS, ulcerative colitis (UC), Rheumatoid Arthritis (RA) and Crohn's disease (CD). (172) They

used random forests method to identify common taxonomic biomarkers specific to IMIDs versus HCs. They found significantly higher abundances of unclassified Erysipelotrichaceae in MS and Intestinibacter in CD, whereas significantly lower abundances of Dialister were found in MS. The ML approach showed best accuracy in classifying disease versus HC for CD (AUC 0.95) and MS (AUC 0.94) followed by RA and UC. The use of ML here highlighted important differentially abundant taxa in different IMIDs which suggest there may be a common element to the aetiology of IMIDs. These different taxa amongst IMIDs may also serve as biomarkers for identifying and diagnosing MS and other IMIDs. The main limitation for this study is the small number of MS patients and the patients being unmatched for both age and sex with HCs. It is also important to note that specific patient data was lacking, including the phenotype, activity and duration of a disease, lifestyles factors such as smoking and diet or the recent use of antibiotics and/or steroids, which are commonly prescribed to the IMIDs group. All these factors are important to consider in future studies as they may have a significant impact on the gut microbiota.

In 2019 Bang et al. suggested that gut microbiome could be classified using ML in six diseases including MS, stroke, juvenile idiopathic arthritis, chronic fatigue syndrome, colorectal cancer and acquired immune deficiency syndrome. (173) They assessed the abundance of microorganisms at five taxonomy levels as features in 696 samples collected from 1079 participants (29 of which had MS) in different studies, to evaluate different prediction models. The models evaluated included SVM, K nearest neighbour (KNN), logistic model tree (LMT) and LogitBoost. They found that using the average accuracies of all four models, the highest classification performance at five taxonomy levels was observed at the genus level (90%). Using 10-fold cross-validation, the performance of classifiers at the genus level reached best accuracy using LogitBoost (94%) and lowest accuracy using KNN (82%). This study showed that selected gut microbiome features could be used to differentiate and diagnose various diseases, including MS, using ML. The main limitations in this study include the small number of MS patients tested and the lack of further validation using external independent cohorts. Obtaining gut microbiome is not performed routinely in clinical

settings and would make the evaluation of this challenging which is an additional limitation.

2.3.1.7 Using electroencephalogram (EEG)

In 2019 Ahmadi et al. used EEG signals to diagnose MS through phase to amplitude coupling in covert visual attention. (174) They used ML methods to detect whether the signals collected indicated MS or not in 5 MS patients (average disease duration of 2.9 years), and 7 healthy controls matched for age and educational level. T-test and Bhattacharyya distance criteria were used to select the relevant and efficient features. LOOCV was used to assess the model validity. Based on phase to amplitude features, this system achieved an accuracy of 91%. This study showed a promising result in ML application using EEG signals during attentional tasks to diagnose MS at an early stage. The main limitations of this study include the sample size and the lack of validation in an external unseen cohort. They also did not include any clinical and radiological features that may be of importance to assess disability and cognition level.

The use of markers other than MRI in MS diagnosis is summarised in Table 10: ML application in MS diagnosis using markers other than MRI. It is interesting to note that there has been a number of studies applying ML in MS diagnosis using different populations, markers and methods of analysis. Many studies have used MRI sequences and tests that are not routinely available in clinical practice and are expensive to perform, which makes further evaluation difficult. It is clear that most studies have used a limited number of participants from a single population and used LOOCV as a method of cross-validation due to the difficulty of assessing methods in external independent cohorts. Therefore, using routinely applied imaging and laboratory tests in the diagnosis of MS through ML and validated in unseen population continues to be an unmet need.

Table 10: ML application in MS diagnosis using markers other than MRI					
	Reference	**Population**	**Markers**	**Methods**	**Accuracy**
1	**Lötsch et al. 2017**	102 MS and 301 HCs	Lipid markers: ceramides, eicosanoids, and lyosophosphatidic acids	Unsupervised method	94%
2	**Ostmeyer et al. 2017**	Training: 11 RRMS and 12 OND. Testing: 60 RRMS and 42 OND	B lymphocytes genetic repertoires	Maximum likelihood optimisation	72%
3	**Forbes et al. 2018**	19 MS, 19 UC, 21 RA, 20 CD and 23 HCs	Gut microbiota	Random forest	MS Vs HCs 94%
4	**Bang et al. 2019**	MS and 6 other diseases	Gut microbiota: five taxonomy levels	LogitBoost SVM Logistic model tree K nearest neighbour	94% 92% 92% 82%
5	**Ahmadi et al. 2019**	MS and non-MS	EEG	Extreme machine learning	91%
6	**Lötsch et al. 2019**	102 MS and 301 HCs	Lipid markers: 8 serum lipid-biomarkers	Unsupervised and supervised, through Bayesian statistics	95%
7	**Pérez del Palomar et al. 2019**	80 RRMS and 180 HCs	OCT: RNFL and GCL+	Decision trees, multilayer perceptron and SVM	Macular area – 97% Macular + Papilla – 95%
8	**Cavaliere et al. 2019**	48 MS and 48 HCs	OCT: the mean of the macular thickness and peripheral areas	SVM	91%
9	**Andersen et al. 2019**	12 MS and 13 HCs	Blood-based metabolites	Supervised ML	6 metabolites achieved accuracy of 80%
10	**Goyal et al. 2019**	From USA: 859 MS and 128 HCs From Russia: 92 MS, 71 HCs	Serum levels of 8 cytokines (IL-1β, IL-2, IL-4, IL-8, IL-10, IL-13, IFN-γ, and TNF-α)	SVM, decision tree, random forest, and neural networks	MS vs HCs, using RF -> 91% RRMS vs. non-RRMS, using RF -> 70%
11	**Ghafouri-Fard et al. 2020**	401 MS and 390 HCs	SNP genotype	Neural network	65%

Abbreviations: ML = machine learning, MS = multiple sclerosis, MRI = magnetic resonance imaging, RRMS = relapsing remitting MS, HC = health control, OND = other immune-related neurological diseases, UC = ulcerative colitis, RA = rheumatoid arthritis, CD = crohn's disease, SVM = support vector machine, EEG = electroencephalogram, RNFL = retinal nerve fibre layer, GCL = ganglion cell layer, OCT = Optical coherence tomography, IL = Interleukins, IFN-γ = Interferon gamma, TNF-α = Tumour necrosis factor α, RF = random forest, SNP = Single nucleotide polymorphisms.

2.3.2 ML application in MS classification:

Having reviewed the literature it is clear that ML has been used to classify MS into different groups and stages. These include radiologically isolated syndrome (RIS), CIS and RRMS, using MRI and clinical parameters. These studies have shown accuracies ranging between 69 – 96%, which indicate that different ML techniques in different studies using different variables can impact the level of accuracy. This highlights that ML algorithms can achieve a high level of accuracy in differentiating subtypes of MS, which can potentially improve patient stratification for future studies looking into treatment response to a variety of therapies. The following section describes studies of ML applications in classifying pwMS.

2.3.2.1 Using MRI

Studies have demonstrated the use of ML application in MS classification using MRI since 2012. These studies are summarised in Table 11: ML application in MS classification using MRI.

Approximately ten years ago Bendfeldt et al. used ML techniques to differentiate between different groups of MS patients using grey matter analysis. (175) They used multivariate SVM analysis to separate grey matter segmentations using 3D-heavily T1w gradient echo images (magnetization-prepared rapid gradient echo 'MP-RAGE') and Transaxial contiguous 3-mm dual-echo T2w images using turbo spin echo in age- and sex-matched groups of MS patients. The groups were divided into early (n=17) or late (n=17) onset MS, low (n=20) or high (n=20) white matter lesion load, and benign (n=13) or non-benign (n=13) MS. This technique showed promising results in detecting MS groups with sensitivity and specificity of 85%, 83%, and 77% respectively after LOOCV. This was one of the first studies into ML application to explore spatial patterns of grey matter segmentations in classifying MS patients, which can potentially be of use in clinical practice. The main limitations include sample size and the lack of validation in an external unseen cohort. A further limitation is the use of non-conventional MRI sequences which would make it difficult to replicate in clinical practice.

In 2014 Crimi et al. proposed a novel characterisation of CIS patients using MRI lesion patterns. (176) They prospectively included 25 CIS patients from multiple centres in France with 3T MRI scans performed within 3 months of disease onset and including T1w, T2w, FLAIR, and pre and post Gadolinium injection sequences. The spatio-temporal lesion patterns identified were used to separate patients in different groups which could potentially undergo a more severe course. An unsupervised clustering algorithm was used and highlighted 3 statistically significant clusters of different lesion patterns at baseline. These clusters also correlated with chronic hypointense lesion volumes by follow up with an R^2 score (a measure of the goodness of fit for the regression model) of 0.90 after LOOCV. This ML methodology helped in classifying MS patients and may lead to better prognostic accuracy at the onset of MS using largely routine clinical MRI sequences. The main limitations of this study were the small sample size and the lack of external validation. It is also worth noting that none of the recruited patients received steroids and the mean EDSS was 1.7 at baseline which indicates a cohort with minimal disability and possibly not reflecting those presenting with spinal cord or brainstem CIS.

In 2015 Weygandt et al. used conventional MRI based ML algorithms to detect early (<12 years) and late (>/= 12 years) onset paediatric MS (EOPMS and LOPMS). (177) They used classification techniques to grey and white matter tissue probability parameters of small brain areas taken from T2w MRI scans of 8 EOPMS, 12 LOPMS and 20 HCs. The paediatric MS group was age-, lesion load- and disease duration-matched to the HCs. It was identified that white matter areas containing lesions differentiated paediatric MS patients from HCs with an accuracy reaching 87%. The prefrontal cortex was found to be specific for EOPMS and had the maximal diagnostic accuracy (77%) after LOOCV using SVM. This study showed that conventional MRI contains good diagnostic features and that grey matter involvement can be used to identify pathognomonic processes specific for EOPMS. The main limitation to this study was the small sample size, which is expected considering the low prevalence of paediatric MS. It is also important to note that there was a significant difference in age between the paediatric MS subgroups which could have a significant impact when using grey and white matter analysis.

In 2016 Kocevar et al. used ML algorithms in combination with graph-derived metrics to differentiate MS clinical profiles. (178) Twelve CIS, 24 RRMS, 24 SPMS, 17 PPMS and 26 age- and sex-matched HCs had 1.5T MRI scans with T1 and diffusion tensor imaging (DTI) sequences. These sequences were used to obtain structural brain connectivity metrics automatically followed by the use of SVM with radical basic function kernel to classify patients. This method showed variable accuracies after 10-fold cross-validation when comparing different groups using binary classification including CIS versus HC (92%), CIS versus RRMS (92%) and RRMS versus PPMS (75%). However, the accuracy dropped (70%) when comparing various subtypes (CIS-RRMS-SPMS). This study showed the possibility of ML to classify MS clinical profiles using T1w and DTI sequences and automatic structural brain connectivity analysis. The main limitations to this study include the small sample size in individual groups and the use of 1.5T MRI scans, which is of less details, and DTI sequences, which are not currently performed in routine clinical practice.

Muthuraman et al. used brain structural network characteristics to differentiate CIS from early RRMS using ML. (179) They identified grey and white matter structural networks in 53 patients (20 CIS and 33 RRMS) and 40 age-matched HCs, using 3T MRI with 3D T1-w MP-RAGE, 3D T2 FLAIR, and DTI sequences. They applied probabilistic tractography and fractional anisotropy maps for white matter and cortical thickness correlation analysis for grey matter. This generated structural connectivity patterns and a network topology analysis was obtained to characterise the network at different community levels (modularity, clustering coefficient, global, and local efficiencies). SVM was used to differentiate between CIS and early RRMS automatically and achieved 97% accuracy using clustering coefficient as a classifier obtained from grey matter. This study demonstrated a ML technique that may help in the early diagnosis of RRMS and may be used to complement the dissemination in time criteria in clinical practice. The main limitations in this study include the small sample size and the fact that RRMS patients were defined according to the 2010 McDonald criteria, which has since been revised. When using the revised 2017 McDonald criteria many of the patients diagnosed with CIS could have been diagnosed as RRMS instead. A further limitation is that DTI and MP-RAGE sequences

are not performed routinely in clinical practice and are therefore difficult to validate externally and subsequently include in the assessment of new patients.

In 2017 Zhong et al. combined structural and functional patterns to classify pwMS according to upper limb motor disability using ML algorithms. (180) Seventy-one age- and sex-matched participants (26 motor function preserved (MP), 25 motor function impaired (MI) and 21 HCs) were included and had 3T MRI with T1w sequences and rsfMRI. Using SVM, discriminant functions indicated regions where grey matter metrics and inter-regional functional connections were identified. The model was able to correctly differentiate and classify upper limb MP and MI. When using grey matter metrics and functional connections measurements the classifier yielded the best accuracy (85%) following LOOCV. This was the first study to identify specific neural substrates which are essential to preserve upper motor function in pwMS, using ML techniques through combining structural and functional features. The main limitation of this study is the relatively small sample size. It is also important to note that a proportion of patients were taking a range of DMTs with various efficacy which could have influenced the pattern of structure and functional connectivity observed.

Ion-Märgineanu et al. used ML techniques to classify the course of MS in combining clinical data with lesion load and magnetic resonance metabolic features. (181) Longitudinal data was obtained from 105 participants (12 CIS, 30 RRMS, 28 SPMS, 17 PPMS and 18 HCs) including clinical measures (Age, EDSS and disease duration) and conventional and spectroscopic MRI scans. The spectroscopic imaging provided the metabolites concentrations data including *N*-acetyl-aspartate (NAA), Choline (Cho), and Creatine (Cre). They analysed 3 features for each metabolic grid by averaging their ratios over good quality voxels. Linear mixed-effects models were generated to establish differences amongst MS groups. Combining clinical features, lesion load and metabolic data, they found significant differences between MS clinical types using 4 different variables: Lesion Load, NAA/Cre, NAA/Cho, and Cho/Cre ratios, through SVM. Using clinical and lesion load gave an accuracy of 71-72% when comparing CIS versus RRMS and CIS versus RRMS and SPMS. Better classification was achieved using clinical and metabolic ratios when comparing RRMS versus PPMS (85%). The highest

accuracy was achieved using clinical, metabolic ratios and lesion load when comparing RRMS versus SPMS (87%). This study was the first to describe the use of clinical data in combination with MRI lesion load and spectroscopic metabolic features to improve discrimination between RRMS and PMS through ML. The main limitations include a small sample size and not taking into account the use of DMTs in the cohort, which could have an impact on results including metabolic rations. It is also difficult to standardise metabolic quantification using MR spectroscopy, as this requires measuring of T1 and T2 relaxation of water for each patient, which would not be possible in clinical practice.

In 2019 Mato-Abad et al. used ML to classify RIS and CIS using multimodal 3T MRI scans with diffusion weighted images to combine cortical thickness, cortical and subcortical grey matter volume and white matter integrity. (182) Seventeen RIS and 17 CIS patients were included. The Naïve Bayes model, using 3 features: the left rostral middle frontal gyrus volume and the fractional anisotropy values in both the right amygdala and right lingual gyrus, achieved the best accuracy (78%) in detecting RIS and CIS. This study highlighted the potential use of ML using MRI to detect early stages of MS; RIS and CIS. The main limitations of this study include the small sample size and the lack of validation in a larger independent cohort.

In 2019 Gonzalez-Campo et al. used ML to differentiate MS patients with fatigue (F-MS) from those without fatigue (nF-MS) or HCs who were age-, gender- and education-matched. (183) They included 49 participants (27 MS patients and 22 HCs), who were assessed for fatigue levels using Modified Fatigue Impact Scale (MFIS) and brain structural and functional connectivity profiles using 3D MRI and rsfMRI. Only F-MS patients showed decreased interoceptive accuracy with decreased grey matter volume and increased functional connectivity in core interoceptive regions, the insula, and the anterior cingulate cortex. These alterations were positively associated with fatigue. Using supervised ML models with SVM, classification between HCs and F-MS patients were achieved with high accuracy (92%). However, the accuracy dropped (69%) when classifying nF-MS and F-MS groups. This study showed a new evidence that fatigue is a potential signature of interoceptive disruptions in MS. The

main limitations for this study include a small sample size and the use of self-reported fatigue scales, which may not be as accurate as other more objective cognitive or motor scales.

Table 11: ML application in MS classification using MRI					
	Reference	**Population**	**Markers**	**Methods**	**Accuracy**
1	**Bendfeldt et al. 2012**	17 Early Vs 17 Late onset MS 20 High Vs 20 low WM lesion load MS 13 Benign Vs 13 non-benign MS	MRI: GM	SVM	85% 83% 77%
2	**Crimi et al. 2014**	25 CIS: severe course pattern	MRI: spatio-temporal lesion patterns	Unsupervised clustering	90%
3	**Weygandt et al. 2015**	8 Early Vs 12 Late onset paediatric MS Vs 20 HCs	MRI: GM and WM	SVM	Prefrontal cortex specific for EOPMS 77%
4	**Kocevar et al. 2016**	12 CIS, 24 RRMS, 24 SPMS, 17 PPMS, 26 HCs	MRI: T1w + DTI	SVM	CIS Vs HC 92%, CIS Vs RRMS 92%, RRMS Vs PPMS 75%, CIS Vs RRMS Vs SPMS 70%
5	**Muthuraman et al. 2016**	20 CIS Vs 33 RRMS	MRI: DTI	SVM	97%
6	**Zhong et al. 2017**	26 motor function preserved Vs 25 motor function impaired Vs 21 HCs	T1w + rsfMRI	SVM	85%
7	**Ion-Märgineanu et al. 2017**	12 CIS, 30 RRMS, 28 SPMS, 17 PPMS, 18 HCs	MRI spectroscopy	SVM	Clinical + lesion loads: CIS Vs RRMS and CIS Vs RRMS and SPMS 72% Clinical + metabolic ratios: RRMS Vs PPMS 85%. Clinical + metabolic ratios + lesion loads: RRMS Vs SPMS 87%
8	**Mato-Abad et al. 2019**	17 RIS Vs 17 CIS	MRI: cortical thickness, GM volume and WM integrity	Naïve Bayes	78%
9	**Gonzalez-Campo et al. 2019**	27 MS (F-MS Vs nF-MS) or 22 HCs	rsfMRI	SVM	F-MS Vs HCs 92% F-MS Vs nF-MS 69%

Abbreviations: ML = machine learning, MS = multiple sclerosis, MRI = magnetic resonance imaging, RRMS = relapsing remitting MS, HC = health control, GM = grey matter, WM = white matter, SVM = support vector machine, CIS = clinically isolated syndrome, EOPMS = early onset paediatric MS, SPMS = secondary progressive MS, PPMS = primary progressive MS, T1w = T1-weighted, DTI = diffusion tensor imaging, rsfMRI = resting state functional MRI, RIS = radiologically isolated syndrome, F-MS = MS patients with fatigue, nF-MS = MS patients without fatigue.

2.3.2.2 Using clinical scales

In 2015 Fiorini et al. used ML to detect 5 different courses of MS including RRMS, PPMS, SPMS, progressive relapsing and benign MS from inexpensive and non-invasive measures such as clinical scales and patient-reported outcomes. (184) The data was collected from 457 MS patients (170 RRMS, 205 SPMS, 68 PPMS, 8 progressive relapsing MS, and 6 benign MS) and 91 variables were described including fatigue, mobility, emotional status, cognition, bladder incontinence and quality of life. The initial data exploration stage, using principle component analysis, found that RRMS patients can be identified and separated from the other clinical courses. Supervised ML techniques, using linear discriminant analysis, were then used to complete feature selection and course classification with modest accuracy of 78%. This was the first study to use clinical scales and patient-reported outcomes to predict RRMS from other courses of MS through ML. The main limitations of this study include the unbalanced data with the majority of patients have RRMS and the lack of clinical data including EDSS that could be of importance in differentiating MS patients of different types or stage.

2.3.2.3 Using genetics

In 2018 Lopez et al. used unsupervised ML technique to identify MS patients' clusters using genomic makeup without providing input parameters a priori. (185) They included 191 patients with MS with over 25,482 single nucleotide polymorphisms (SNPs). This method was able to group 96.3% of patients in the same clusters when both the complete dataset and the different data subsets were used, with accuracy reaching 96% after a 10-fold cross-validation. The identification of distinct genetic subtypes of patients shows the potential use of this method to lead personalised medicine of heterogenous conditions with heritable components, such as MS. This finding is important to explore and validate in a much larger, independent dataset.

Once again, this section has demonstrated that routine clinical data used in the field of ML to classify MS has been limited. Within these limited studies, most have focused on one aspect, such as MRI, clinical characteristics, or genetics, rather than

a combination or a holistic approach. Most studies have lacked the inclusion of large cohorts from different populations, e.g. participants from multiple centres, and validation in an external and independent dataset. This is ultimately much needed and would be more reliable in the field of ML when identifying clinically meaningful outcomes in the classification and management of MS.

2.3.3 ML application in MS prediction

The application of ML in MS prediction has been investigated for nearly 15 years. Most studies have focused on two aspects including predicting the possible progression of disability and the response to the rapidly evolving therapies in order to help clinicians to choose the best treatment for MS patients. In both cases, ML algorithms have shown variable accuracies across different studies. For predicting disease progression, ML methods have reported accuracies ranging 61-92%, whereas ML methods in predicting treatment response have reported accuracies ranging 68-89%. The following are summaries of the reported studies in the literature, as demonstrated in Table 12: ML application in MS prediction.

2.3.3.1 Predicting conversion of CIS to MS

In 2008 Corvol et al. used supervised ML to build predictive models of CIS conversion to MS using microarrays studying gene expression in naïve CD4+ T cells. (186) Including 66 participants (37 CIS patients and 29 HCs matched for age and gender), they generated 1,718 genes which showed clear segregation between CIS and HCs samples, using a hierarchical clustering with 99.8% accuracy. Furthermore, 975 genes showed expressions that differentiated CIS patients into 4 subgroups with >80% accuracy using a supervised ML algorithm, Integrated Bayesian Inference System. One subset showed 108 genes that further differentiated CIS patients with 92% of them converted to MS. The robustness of this group over time was measured using SVM with 10-fold LOOCV, with 100% accuracy at baseline. However, this became a less powerful predictor after 12 months with 86% accuracy. This is one of the first studies to show that ML methods can potentially be used to classify CIS patients into

different groups and differentiate them from HCs. The main limitations for this study include a limited sample size and the lack of clinical information, such as the OCBs status which are important in the diagnosis of MS for CIS patients, and radiological data.

No further studies were noted until 2014 when Wottschel et al. showed that ML algorithms can predict a second relapse leading to the diagnosis of MS in CIS patients. (187) They used conventional MRI (T2w and proton density (PD)) features (number of lesions) and clinical characteristics (gender, age, EDSS, clinical phenotype at onset) of 74 CIS patients collected retrospectively. Using SVM they predicted conversion to MS with 71% accuracy at 1 year and 68% accuracy at 3 years after LOOCV. It was noted that the use of a combination of features always predicted outcome better than any single feature. This study highlighted that ML has the potential to predict the conversion of CIS to MS in individuals using baseline clinical features and conventional MRI scans and can therefore be potentially used to support clinical practice. The main limitations for this study include a small sample size and the lack of additional radiological (such as brain volumes) and clinical (such as the presence of OCBs) features. It is also important to validate these results in an independent and larger multicentre cohort.

In 2019 Zhang et al. predicted the conversion of CIS to MS using ML methods based on baseline MRI images (3-D FLAIR and T1w). (188) They evaluated 84 CIS patients, from a prospective observational single centre cohort, with a follow up of at least 3 years. They developed brain lesion masks by a computer assisted manual segmentation and produced automatic segmentations using Lesion Segmentation Toolbox for SPM (Statistical Parametric Mapping) to assess different segmentation methods. They predicted the conversion of CIS to MS, or non-conversion, using random forest methods with accuracy reaching 85% at best, after 3-fold cross-validation, based on shape features produced by the segmentation masks. This study showed that shape parameters of lesions can predict the conversion of CIS to MS at an early stage with good accuracy, which can be of significant importance in starting early treatment. The main limitations to this study include a small sample size and

the fact that clinical features such as demographics and EDSS were not included in the predicting models, which would be of importance in clinical practice. It is also worth mentioning that basing prediction models on lesion shape can be difficult in 2D images in case of small MS lesions.

Wottschel et al. used another ML technique in 2019 to predict relapses in the short term for CIS patients, and converting to MS, using MRI brain at baseline with T1w sequences. (189) They evaluated 400 CIS patients retrospectively, from 6 MAGNIMS (Magnetic Resonance Imaging in MS) centres, with clinical (age, sex, EDSS, clinical phenotype of CIS onset) and brain MRI characteristics at baseline. For each patient they assessed MRI brain features including grey matter probability, white matter lesion load, cortical thickness, and volume of specific cortical and white matter regions. They used recursive feature elimination and weight averaging with SVM to distinguish CIS converters versus non-converters at one year follow up. SVM showed that the most relevant MRI features in predicting a further relapse were white matter lesions, grey matter probability in the thalamus and the precuneus or cortical thickness in the cuneus and inferior temporal gyrus. The model's accuracy in distinguishing both groups across centres using LOOCV varied between 73-93% depending on the site being assessed. The higher accuracy sites were associated with a lower number of data sets, which possibly indicates overfitting or selection bias. Understandably, due to the heterogeneity of the data, assessing all centres together provided a lower accuracy of 70.8%. The main limitations of this study include the retrospective nature of the data collected and the lack of imaging harmonisation. The accuracy of this model may improve with the addition of other factors such as genetics and environmental factors which could be informative for individualised prognosis.

2.3.3.2 Predicting conversion of RRMS to SPMS

There have not been many ML studies in this field, however in 2020 Seccia et al. assessed the impact of routine clinical records using ML methods in predicting the conversion of RRMS to SPMS. (190) They interrogated 1624 outpatients' clinical records, including 207 SPMS patients, and divided predictors at 6, 12 and 24 months.

They compared 4 ML methods (SVM, RF, K-Nearest Neighbours and AdaBoost) in the visit-oriented setting or a Recurrent Neural Network model for history-oriented setting. The classifiers were assessed by Recall (sensitivity) and Precision (positive predictive value). The Recall of the visit-oriented setting reached 70-100%, whereas the Precision was low (5-10%). In comparison the history-oriented setting reached Recall of 67% and Precision of 42% at 2 years. For the first time this study showed that clinical data can be used to predict the course of MS which could be useful to guide therapy for those at increased risk of conversion. The main limitation of this study is the unbalanced data with a small number of patients converting to SPMS in the history-oriented setting, which may have compromised the performance of the models. It is also important to mention that such findings would need to be tested and compared in an external independent large cohort.

2.3.3.3 Predicting disability and progression

This section of ML has been of interest over the last five years due to its importance of identifying patients at high risk of increased disability, which could potentially be targeted for future therapeutic trials.

In 2017 Zhao et al. used ML algorithms to predict disability progression in MS patients. (191) They analysed 1693 patients from the CLIMB study and classified patients as disease progression with increased EDSS by ≥1.5 or not at five years follow up after baseline assessment. They collected data, including demographics (age, sex, ethnicity, family history of MS, smoking), clinical (EDSS) and MRI findings (whole brain T2 lesion volume (T2LV) and normalised whole brain volume), at year one and two. They subsequently applied SVM and logistic regression to predict the EDSS score at five years. They found that, using SVM, baseline information alone had little predictive value. However, adding data at one year improved the overall sensitivity (62%) and specificity (65%) in predicting EDSS progressing patients. Further improvement was noticed with the addition of MRI data at one year with sensitivity (71%) and specificity (68%) after 10-fold cross-validation. SVM performed better than logistic regression in most cases. Family history of MS, ethnicity and brain parenchymal fraction had the highest prediction for the non-progressing group.

Whereas T2LV were the highest predictor of EDSS worsening. This study highlighted a promising step towards the use of ML in predicting MS disease course using clinical and imaging data which can help in potentially selecting more aggressive therapies for the most suitable patients at an early stage. One of the limitations of this study includes the use of EDSS as a scale of progression which is less sensitive to upper limbs function and cognitive decline. Another limitation includes not using DMTs as predictors which could have effects on disease course.

In 2018 Kiiski et al. used ML to predict cognitive functioning and processing speed at the individual level. (192) They used EEG over a 2 year follow up period in 35 MS patients versus 43 age-matched HCs. Recording event-related potentials (ERPs) using EEG and a neuropsychological battery at different intervals gave an objective assessment of cognitive function and processing speed. They used penalised linear regression modelling and internal 10-fold cross-validation using all spatio-temporal ERP to predict cognitive functioning and information processing speed at baseline and a year later. They compared EEG and non-EEG models where non-EEG variables included age, handedness, occupational status and years of education. They found that ERPs during visual oddball tasks predicted cognitive functioning and information processing speed best (EEG only model $r = 0.27$, EEG + non-EEG model $r = 0.37$). However, ERPs during auditory tasks were not predictive. This study showed that EEG, although not commonly used in the field of MS, and ML show an exciting outlook in predicting outcomes, which can be of use in individualised management plans considering early cognitive impairment is associated with poorer outcomes. (193) The main limitations of this study include the small sample size and the lack of external validation.

In 2019 Law et al. used ML to predict short term disability progression in SPMS patients. (194) They evaluated 485 SPMS patients as progressors if a 6-month sustained EDSS worsening was noted. They included the following features; age, sex, disease duration, T2 lesion volume and multiple sclerosis functional composite component scores. They built models using decision tree (DT) and compared to logistic regression and SVM for predicting disease disability progression. DT was

found to be the best model for prediction with accuracy reaching 62% after internal 10-fold cross-validation. Although showing poor level of accuracy, this was the first study to use ML techniques to predict SPMS patients with highest risk of short-term progression and can potentially help SPMS investigators to select such cohorts for potential therapies in the future. One of the main limitations of this study includes the imbalance between those who progressed versus those who didn't. The other limitation includes the use of EDSS as a measure of progression which fails to account for cognition.

In 2020 a genetic model of MS severity to predict future disability was first used by Jackson et al. (195) They evaluated genetic modifiers of MS severity in training (n=205) and validation (n=94) cohorts acquired prospectively. They assessed 113 genetic variants related to MS severity and used an MS disease severity scale (MS-DSS). Random forest was used as a ML method to formulate a single genetic model of MS severity (GeM-MSS). They found that GeM-MSS correlated significantly with MS-DSS (r = 0.214; p = 0.043) when assessed in a validation group, although of much lower strength compared to the training cohort (r = 0.969, $p = 8.76 \times 10^{-125}$), that was not involved in the modelling phase. This was the first independently validated study to apply ML in predicting MS severity and future disability using genetics. This has the potential to find therapeutic targets for stopping disability progression in MS. However, it would be essential to validate this method in a much larger cohort with more clinical dataset, such as EDSS.

Using motor evoked potential (MEP) time series, Yperman et al. predicted MS progression through ML. (196) They included 419 MS patients and generated a total of 6219 MEP visits. They used random forest (RF) and logistic regression classifiers to predict disease progression at 2 years. RF achieved the best performance in predicting disease progression with accuracy reaching 75%. This study highlighted the use of MEP time series in predicting MS disability progression through ML. However, this requires further validation in larger and multi-centre cohorts in the future. The prediction performance may also improve when adding further features such as MRI metrics and the use of DMTs which are important in clinical practice.

Meanwhile, Roca et al. built a ML algorithm that predicts EDSS score in pwMS at 2 years based on age, sex and FLAIR MRI scans only. (197) They combined convolutional neural network (CNN) and classical ML predictors based on random forest regressors and manifold learning trained using the location of lesion load in relation to white matter tracts. They included 971 MS patients for the training dataset (including 10% used for validation) and tested 475 patients (with no known EDSS score). The resulting algorithm predicted EDSS score with a mean square error of 2.2 in the validation cohort and a mean square error of 3 (mean EDSS error = 1.7) in the test cohort. This method gave promising results in predicting EDSS at 2 years using MRI FLAIR and basic demographics. This however should be further validated in a larger external test group as it has the potential to be used for future prediction of disability and subsequently DMT evaluation.

Table 12: ML application in MS prediction						
	Reference	**Population**	**Prediction**	**Markers**	**Methods**	**Accuracy**
1	**Corvol et al. 2008**	37 CIS + 29 HCs	CIS -> MS	gene expression in naïve CD4+ T cells	Hierarchical clustering and Bayesian statistics	92%
2	**Wottschel et al. 2014**	74 CIS	CIS -> MS	MRI + clinical + sociodemographic features	SVM	71% & 68% at 1 & 3 years respectively
3	**Zhao et al. 2017**	1693 MS	EDSS progression at 5 years	Demographics, clinical and MRI measures	SVM	86%
4	**Kiiski et al. 2018**	35 MS + 43 HCs	Cognitive functioning + processing speed	EEG and a neuropsychological battery over a 2 year follow up	Linear regression	EEG only model r = 0.27. EEG + non-EEG model r = 0.37
5	**Zhang et al. 2019**	84 CIS	CIS -> MS	MRI (characteristics of lesions)	Lesion Segmentation Toolbox for SPM	85%
6	**Wottschel et al. 2019**	400 CIS	CIS -> MS	MRI (WM lesion load, cortical thickness and volume of cortical and WM regions)	SVM	73-93%
7	**Law et al. 2019**	485 SPMS	EDSS progression at 6 months	age, sex, disease duration, T2 lesion volume and MS functional composite component scores	Decision tree, logistic regression, SVM	DT - 62%
8	**Seccia et al. 2020**	207 SPMS	RRMS -> SPMS	MRI + demographic, clinical, and treatments	SVM, Random forest, K-Nearest Neighbours, AdaBoost, and Neural network	History-oriented setting PPV 42% at 2 years
9	**Jackson et al. 2020**	299 MS	MS disability progression	Genetic modifiers of MS severity: 113 genetic variants	Random forest	Genetic model correlates with MS-DSS (r=0.214; p = 0.043)
10	**Yperman et al. 2020**	419 MS	MS progression at 2 years	Motor evoked potential (MEP)	Random forest and logistic regression	RF – 75%
11	**Roca et al. 2020**	971 MS	EDSS score	Age, sex and MRI	Random Forest	Mean EDSS error = 1.7

Abbreviations: ML = machine learning, MS = multiple sclerosis, CIS = clinically isolated syndrome, HC = health control, MRI = magnetic resonance imaging, SVM = support vector machine, EDSS = expanded disability status scale, EEG = electroencephalogram, SPM = statistical parametric mapping, WM = white matter, SPMS = secondary progressive MS, RRMS = relapsing remitting MS, PPV = positive predictive value, MS-DSS = MS disease severity scale.

2.3.3.4 Predicting treatment response

Predicting treatment response in MS using ML has been investigated for over five years. The studies available in the literature are summarised in Table 13: Prediction of treatment response in MS.

Initially Baranzini et al. looked into prognostic biomarkers of response to subcutaneous beta interferon (IFN-β) therapy in RRMS patients using ML. (198) They evaluated 39 genetic biomarkers from 155 RRMS patients in the IMPROVE study with data including clinical assessment, blood samples and MRI scans. Random forest classification algorithm was used to evaluate combinations of baseline biomarkers that correlate with treatment response over time. Treatment response was defined by absence of new MRI lesions, relapses, and increase in EDSS over 1 year. They found that a combination of 3 genes at baseline was able to identify patients with poor therapeutic response with accuracies reaching up to 80% (with CASP2/IL10/12Rb1 genes combination). This was the first study to look into predicting treatment response through ML using genetic biomarkers which can potentially be of prognostic value in future clinical practice. One of the main limitations of the study include the disproportion between responders and non-responders in a small sample size, which could cause direct consequences on the performance of the predictors.

In 2017 Kalincik et al. used ML to predict treatment response to 7 DMTs in the large Australian MSBase by collecting a variety of data including 27 demographic, clinical and paraclinical predictors. (199) They analysed the response to treatment by assessing relapse frequency, disability progression, conversion to SPMS and treatment discontinuation. Generalised linear and multivariable survival models were implemented with principal component analysis (PCA). The main modifiers in the training data were age, disease duration, EDSS, previous relapses and previous treatment. For injectable therapies they found that higher disability worsening over time was associated with higher EDSS at therapy initiation and the use of previous DMTs. For other higher efficacy therapies, such as Fingolimod, Mitoxantrone and Natalizumab, it was predominantly associated with lower relapse activity prior to therapy initiation. The incidence of future relapses was associated with age and

relapse activity prior to starting therapy. The accuracy of the predictive models was above 80% for predicting relapses and disability outcome at year one, but reduced at year 2-4. The models had low accuracy (<50%) in predicting future conversion to SPMS and treatment discontinuation. Similar results were observed when performing external validation. This was one of the first studies to use real-world large data highlighting the importance of using clinical and paraclinical data. One of the main limitations of the study include the lack of detailed imaging data, which is essential in clinical practice and the potential future use to identify the right patients for the right therapy at the right time to maximise outcomes for pwMS. Another limitation includes the small sample size of some of the DMT cohorts including cladribine, alemtuzumab, teriflunomide and dimethyl fumarate, which affected the development of informative predictive models.

In 2019 Kanber et al used ML to compare the ability of conventional models to identify imaging-based response to natalizumab against high-dimensional models integrating a variety of imaging factors. (200) They included 124 natalizumab treated RRMS patients with routine standard of care MRI scans including T1w and FLAIR sequences. They used fully-automated image analysis to obtain 144 regional, longitudinal trajectories of pre- and post- therapy changes in brain volume and disconnection. They applied ML to build and evaluate low- and high- dimensional models of the correlation between natalizumab and the trajectories of change. Using AUC as a measure of accuracy, high-dimensional models were much better at detecting treatment response when compared to low-dimensional models (0.890 [95% CI = 0.885–0.895] vs. 0.686 [95% CI = 0.679–0.693], P < 0.01]). This highlighted that high-dimensional models extracted from routine clinical imaging can significantly improve detection of treatment response which can potentially lead to a more accurate individual prediction. The main limitation of this study includes the use of a cohort of MS patients on one DMT. Future assessments of this method on other cohorts and on other DMTs would be of significant importance.

Predicting response to natalizumab was also investigated through ML by Fagone et al. using a subset of genes of CD4+ T cells in RRMS patients prior to starting therapy.

(201) They included Swedish RRMS patients who were classified as low responders (LR, n=6) and high responders (HR, n=6) and were matched for gender, age, EDSS and disease duration. The parameter used to evaluate responsiveness was the relapse rate. All samples for gene expression analysis were obtained prior to starting natalizumab. Principle Component Analysis (PCA) and Hierarchical Clustering (HCL) were used to evaluate predictors. A total of 17 predictors of 45 probes were found following analysis of the transcriptomic differences between CD4+ T cells from the HR and LR groups. HCL and PCA were able to use the 17 predictors to accurately differentiate LR and HR patients with 89% accuracy. This was the first study to identify genes that can differentiate MS patients that are high or low responders in a small natalizumab cohort. Although this finding requires further investigation and validation in a larger cohort, it highlights that genetic biomarkers can potentially be used to predict treatment resistant MS patients which would lead to better understanding of therapeutic targets and outcomes.

In 2020 Ebrahimkhani et al. used ML to predict MS activity in MS patients taking fingolimod using serum exosome microRNAs. (202) MS activity was assessed with the presence of gadolinium-enhanced MRI brain at baseline, 6, and 12 months post therapy initiation. Next-generation sequencing was used to profile serum exosome microRNAs collected at baseline and 6 months from 29 RRMS patients. Using logistic regression, 15 microRNAs were identified from post-fingolimod therapy patients which were differentially expressed in patients with active versus non-active MS. Random forest was used to predict combinations of multiple microRNAs as multivariate signatures of MS activity. The prediction of response using single microRNA was modest with an average of 77%. However, using combinations of 2-3 microRNAs reached greater accuracy (90%) in distinguishing active versus non-active patients. This ML study emphasised the importance of serum exosome microRNAs as potential biomarkers of disease activity in RRMS patient. This requires further evaluation in a larger independent cohort.

Finally, Jin et al. used ML to predict response to IFN-β therapy in MS. (203) They generated a feature selection method from gene-to-gene interactions. They used this on a longitudinal microarray dataset from 25 German RRMS patients and assessed the treatment response over 2 years by identifying the gene pairs that correspond to responders versus non-responders. Using SVM, it reached a predictive accuracy of 81% using 41 gene signatures. This study has shown the potential of ML in predicting response to treatment using genetics which is gaining more interest towards personalised response to DMTs in MS. The main limitation of this study includes the sample size and therefore a larger gene expression study is warranted to assess these findings.

Table 13: Prediction of treatment response in MS						
	Reference	**Population**	**Prediction**	**Markers**	**Method**	**Accuracy**
1	**Baranzini et al. 2015**	155 RRMS – IFN-β (RCT)	Clinical and radiological response to treatment	39 genetic biomarkers: Combination of 3 genes at baseline	Random forest	63-80%
2	**Kalincik et al. 2017**	8513 MS; On 7 DMTs in Australian MSBase	Relapse frequency, disability progression, conversion to SPMS, DMT discontinuation	27 demographic and clinical predictors	Generalised linear and multivariable survival models with principle component analysis	Relapses + disability at year 1 – 80%, lower at 2-4 years. conversion to SPMS & DMT discontinuation - <50%
3	**Kanber et al. 2019**	124 RRMS – Natalizumab	Imaging treatment response	Routine clinical MRI brain: High and low dimensional features	SVM and extreme randomised trees	High dimensional = 89% Low dimensional = 68%
4	**Fagone et al. 2019**	12 RRMS - Natalizumab	Low responder Vs high responder	Subset of genes of CD4+ T cells + MRI and relapse rate data -> 17 predictors generated	PCA and Hierarchical Clustering	89%
5	**Ebrahimk-hani et al. 2020**	29 RRMS - Fingolimod	Active Vs non-active MS: Gd-enhanced MRI brain lesion at 6 and 12 months	15 serum exosome microRNAs identified	Random forest	Combinations of 2-3 microRNAs - 90%
6	**Jin et al. 2020**	25 RRMS - IFN-β	Responders Vs non-responders over 2 years	Gene-to-gene interactions: 41 gene signatures	SVM	81%

Abbreviations: MS = multiple sclerosis, RRMS = relapsing remitting MS, RCT = randomised controlled trial, DMTs = disease modifying therapies, SPMS = secondary progressive MS, SVM = support vector machine, PCA = principal component analysis, Gd = Gadolinium, IFN-β = interferon beta.

2.4 Conclusions

In summary, ML application in MS has been used for diagnosis, classification, disease progression and treatment response for nearly two decades. Different aspects have been utilised in these applications including clinical, radiological, immunological and genetics among others, with different techniques reaching a high level of accuracy in many cases.

It is important to note that comparisons between the different categories should be considered with caution due to the different techniques used, the number of variables included and the precision measures used, which are different amongst most studies. It is also clear that the studies reviewed above have mostly been novel, and despite showing promising results, they have often included a limited number of patients and in most cases require further testing and validation in independent and much larger external cohorts.

Overall, we have highlighted the enormous applications of ML in the field of MS and new avenues of research are continuing to be explored, particularly in the field of treatment response to the various pharmacological therapies that currently exist. Although there is no substitute for ongoing randomised controlled trials in providing the best level of evidence, the application of the emerging vast real-world data is likely to help in developing ML applications in the future.

Although one of the limitations to our review is that it is not systematic, we have conducted an extensive, detailed and step-by-step review of the literature using one of the most commonly used search engines, PubMed, by clinicians. We have tried to capture as many of the studies as possible to demonstrate the wide use of ML methods in MS. However, due to the use of one search engine, some studies may have been missed. We may have introduced bias as our studies were screened and selected by one reviewer, myself. In our methodology we did not include a specific method of analysis of results and sensitivity, but we included all studies and reported all outcomes, and provided source of possible bias and limitations.

Although recent studies have created a framework using possible predictors to identify treatment response, it is worth noting that these studies have focused their predictions at a population level rather than on the individual level. This is an area of increasing interest and will be the focus of ML in MS in the coming years considering increasing popularity of large real-world data. The importance of this is becoming apparent to investigators, from the promising evidence that already exists that potentially will lead the way towards personalised medicine which will help healthcare systems to implement ML tools in clinical practice.

The next chapters of this thesis will focus on our work towards prediction of treatment responses in MS using mostly retrospective data collected from the electronic records at Queen Square Multiple Sclerosis Centre at the National Hospital for Neurology and Neurosurgery.

Chapter 3. Identifying predictors of MS treatment response in a single-centre cohort

3.1 Introduction

Highly effective DMTs are becoming widely available in the MS field which has led to the evolution of achievable treatment goals. No evidence of disease activity (NEDA) is widely accepted by MS neurologists as a comprehensive assessment of treatment outcome. NEDA-3 combines MRI activity and clinical progression (204, 205), as described in Table 14: NEDA-3: No evidence of disease activity parameters **in MS**. Brain atrophy has also been suggested to be included in the assessment and monitoring of pwMS on DMT (NEDA-4) (206), however this has not been part of routine clinical practice yet. NEDA in the remaining of this thesis will imply NEDA-3.

Table 14: NEDA-3: No evidence of disease activity parameters in MS	
Clinical	1. No relapses
	2. No disability progression (EDSS progression)
MRI	3. No MRI activity (new or enlarging T2 lesions or Gd-enhancing lesions)

Predicting response to treatment in pwMS and achieving NEDA is much needed for many reasons. There are many different DMTs available to reduce clinical relapses and they have different efficacy and safety profiles. Also, individual patients respond to DMTs very differently and up to 60% of them have a new relapse within 2 years which may lead to an accumulation of disability and poor compliance to therapy. (207)

The disease course and clinical outcomes in MS have been associated with clinical and radiological features. It was found that patients taking interferons, who continued to have relapses and MRI activity during the first year of therapy, have a high risk of disability. (208) Spinal cord presentation is fundamental as a diagnostic and a prognostic factor in MS as acquiring spinal cord lesions as an early presentation or during the disease course is associated with an increased risk of disability

accumulation and therefore progressing to secondary progressive MS. (209) It has also been shown that spinal cord lesions on MRI at onset are associated with a more aggressive MS and disability progression. (7)

Recently, Sormani et al. predicted the risk of clinical disease activity and disability progression in patients treated with subcutaneous interferon beta-1a over a 14-year period using the MAGNIMS score at year 1. (210) The MAGNIMS score (0-2) considers clinical relapses and MRI brain activity (the number of new T2 lesions). (208) They found that those with a MAGNIMS score of 2 had a significantly higher risk of having clinical disease activity (relapses and EDSS worsening) compared to those with a score of 1, who in turn had a significantly higher risk of a clinical disease activity compared to those with a score of 0. Similarly, this trend was also observed on assessing the risk of EDSS progression. One of the main limitations to this study is the lack of MRI data over the period of evaluation hence NEDA was not feasible in this study. It would also be important to identify predictors at baseline for better early decision-making regarding management and long-term outcomes.

Early brain and spinal cord imaging predictors of long-term (15 years) outcomes in RRMS including physical disability, cognitive performance and conversion to SPMS were investigated by Brownlee et al. (7) They found that baseline gadolinium enhancing lesions and spinal cord lesions were independently associated with conversion to SPMS at 15 years. Those measures showed a consistent association with EDSS at 15 years. Also, baseline gadolinium enhancing lesions were associated with worsening cognitive performance at 15 years. The main challenge in using these factors is that MRI with contrast and spinal cord MRI scans are not routinely performed in clinical practice in the NHS, unless clinically indicated or for a diagnostic purpose.

Demographics, clinical and radiological predictors of 30-year outcomes in a group of 120 CIS patients were investigated by Chung et al. (8) Sex and disease duration showed no association. Older patients at onset of CIS and an infratentorial (IT) presentation (compared to optic neuritis or spinal cord) were associated with a

greater risk of MS-related mortality. The change in EDSS over the first 5 years was also seen the most amongst those with an IT presentation. However, baseline EDSS scores were not associated with a 30-year outcome, including developing SPMS. Total lesions count and the presence of IT lesions at baseline were the early MRI predictors of 30-year EDSS of >3.5. The presence of IT lesions and deep white matter lesions at year 1 were also strong predictors of 30-year EDSS of >3.5 and developing SPMS. This is similar to previous reports of IT lesions linked with less favourable outcomes in MS patients. (211) Changes in lesion count and presence of new lesions were not predictive most likely due to missing data. The main limitation of this study is the fact that this cohort were recruited in the 1980s and MRI was a new technique with low image quality, limited to T1w sequences without contrast and therefore the assessment of new and/or active lesions was limited.

Considering the value of clinical and radiological predictors, it is important to consider other patient aspects that may shape the individual response to treatment that can guide DMT selection. It is also important to note that DMTs have different efficacy and, although there has been no head to head comparisons particularly for similar efficacy therapies, higher efficacy therapies are expected to be more associated with NEDA. In this study we explore all the variables including demographics, clinical and imaging data available to the MS neurologist at our centre. Using frequentist statistics, multiple logistic regression, we aim to explore the variables that are significantly associated with NEDA at 1 and 2 years in five of our cohorts separately in this chapter.

3.2 Methods

Initially, the focus was on studying a retrospective cohort of MS patients who have had DMTs at Queen Square Multiple Sclerosis Centre (QSMSC) at the NHNN. Another prospective cohort as part of the PITMS (Predicting Individual Treatment responses towards personalised medicine in Multiple Sclerosis) study was also included. PITMS is an observational prospective study which aims to recruit patients with MS initiating a DMT in the National Health Service (NHS). The plan was to investigate the baseline clinical and MRI characteristics of DMT responders and non-responders. A baseline was defined as the time of starting the index DMT. A responder was defined as a patient who achieved NEDA in the short and medium-term (12-24 months) of commencing a DMT. We conducted and reported this study according to STROBE (strengthening the reporting of observational studies in epidemiology) and a checklist has been provided in the appendix at the end of the thesis. This work had the ethical approval from the research ethics committee as per the NIHR PITMS study: 19/WA/0157. PITMS was funded by the National Institute for Health Research (NIHR) and the UK MS Society.

3.2.1 Subjects

The retrospective cohorts were identified through the MS clinical nurse specialist (CNS) and pharmacist databases at QSMSC. I performed the data extraction through the electronic patient health record system (EPIC), which has been used at the NHNN since 2019. The patients' records go as far back as 2008 with further historic letters scanned and filed into EPIC. Identifying patients through the pharmacy and MS speciality nurse databases therefore reduced selection bias. The data was collected using MS team letters, DMT MDT outcomes and notes from MS consultants, registrars and specialist nurses. Physiotherapy and rehabilitation assessments were helpful in working out EDSS scores when possible. Pharmacy notes were evaluated for medication compliance and adverse events. Using the electronic records have optimised our collection of data and minimised sampling bias. Patient were followed up on annual basis from DMT initiation (baseline), as per NHS routine clinical care.

Considering the large number of DMTs used in the last three decades, we identified four retrospective cohorts to analyse including two first line moderately effective therapies, glatiramer acetate (GA) and dimethyl fumarate (DMF), a second line more effective therapy, fingolimod, and a highly effective therapy for very active MS, natalizumab. These therapies were chosen because they have been used for over 10 years and are still being used routinely today.

A fifth, prospective, cohort was also included from the PITMS study. Data of RRMS patients who commenced ocrelizumab was collected considering it is a high efficacy DMT, one of the latest approved DMTs, frequently prescribed nowadays due to its effectiveness, relatively low risk profile, convenient administration and good availability, with less strict criteria to initiate. The data for the prospective cohort was stored anonymously in REDCap, which is a secure web platform for building and managing online databases and surveys. The data was stored securely in REDCap, accessed only by authorised personnel, and automatically exported for seamless data downloads to excel and common statistical packages, such as R.

3.2.2 Inclusion/Exclusion criteria

We included any RRMS patient, aged ≥ 16 years old, who commenced the specified DMTs at QSMSC from 2002 (as far as the records can be accessed) and last followed up in February 2022. Patients on DMTs for PPMS or SPMS were excluded.

3.2.3 Analysis using combined multiple logistic regression

In this chapter combined multiple logistic regression (LR) modelling is used to look for clinical predictors of NEDA in each of our cohorts.

LR predicts a binomial outcome from one or more predictor variables. It assessed the impact of each variable observed on the outcome of interest and at the same time, so the resulting population-level effect estimates (odds-ratios) of each predictor variable are adjusted for each other. This is also known as the frequentist method.

One model was designed, combining all available clinical variables, for each cohort to identify baseline predictors to the binomial outcome of NEDA at each time point. All analysis was performed using RStudio v1.4.1106 (2021). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/).

All binary categorical variables were represented by single terms where 1 = presence and 0 = absence of the feature. DMTs, ethnicity and clinical phenotype at MS onset were all assigned into five categories each and so they were assigned as five individual categorical variables (represented as five terms for each variable). Only patients who commenced therapy before September 2021 and September 2020 were included in the NEDA analysis at 12 and 24 months respectively. Those who stopped therapy in less than 6 months and less than 18 months (for reasons other than disease activity) were not included in the NEDA prediction analysis at 1 and 2 years respectively.

3.2.4 Missing Data

Collecting data retrospectively from an electronic health record inevitably resulted in some missing data. Similarly, for the prospective cohort, due to the COVID-19 pandemic and limited hospital visits and patient exposure, missing data was also unavoidable. Restricting analysis to only those with fully available datasets can cause bias as well as reducing study power.

Multiple imputation was used to impute the missing data assuming that it was missing at random conditional on the other covariates measured. This was performed in the multiple imputation by chained equations (MICE) package in R (version 3.12.0). (212) This MICE package helps in imputing missing values with plausible data values drawn from a distribution specifically designed for each missing data point. The mice function takes care of the imputing process once the code is written correctly. We used the following code: mice (DMT, m=n, seed = 123), where “DMT” is the name of the therapy analysed, “m” is the number of imputations used and seed is the random parameter which is useful for reproducibility.

3.3 Clinical and MRI Data

All cohorts had the following data collected at baseline (at time of starting the index DMT);

- Demographics: sex, ethnic background and age at MS onset
- Clinical history: the clinical phenotype at MS onset, previously used DMTs, disease duration prior to DMT initiation, number of relapses in the previous 12 months, comorbidities (the presence of at least one past medical history) at baseline
- Examination: EDSS at baseline
- Investigations: the presence of unmatched OCBs in CSF, genetics (including the presence of one/two alleles to the gene coding for HLA-DRB1*15)
- Imaging: number of new MS lesions on MRI brain reported by neuro-radiologists. Baseline MRI features were collected from scans taken in the first 6 months of taking DMTs, unless scans from this time period were unavailable, in which case scans taken in the 6 months prior to initiating therapy were utilised. New MS lesions were defined as new T2 and/or Gadolinium enhancing MS lesions identified when compared to previous scans, performed in the previous 1-3 years.

The following information was collected at one and two years from starting therapy;

- Number of relapses; a relapse is defined by new/worsening neurological symptoms lasting more than 24 hours in the absence of another cause such as an infection.
- Number of new MS lesions on the brain MRI scan compared to baseline scan
- EDSS scores
- Whether NEDA was achieved; defined as no evidence of relapses, no new MRI brain lesions and no significant increase to EDSS. NEDA at 12 and 24 months are demonstrated in Table 15: No evidence of disease activity (NEDA) components. A significant increase in EDSS was defined as an

increase of EDSS by 1.5 or more if previous EDSS was 0, increase of EDSS by 1.0 or more if EDSS was between 1.0 and 5.0, and an increase of EDSS by 0.5 if previous EDSS was 5.5 or more.

Table 15: No evidence of disease activity (NEDA) components		
NEDA components	**NEDA at 12 months**	**NEDA at 24 months**
1. No. of Relapses in previous 12 months	0	1
2. Any worsening of EDSS compared to previous 12 months?	No	No
3. No. of new MRI brain lesions compared to previous MRI brain	0	0
NEDA? Yes or No	Yes	No

3.4 Results

Having collected data for all five cohorts, a total of 1986 patients have been evaluated. The following sections will demonstrate details of our cohorts in terms of demographics, clinical and radiological features, and the outcome of NEDA at 12 and 24 months.

3.4.1 Analysing demographics, clinical features and NEDA outcomes

3.4.1.1 Dimethyl fumarate (DMF)

DMF is a first line therapy, which had 670 patients initiating treatment between January 2012 and March 2020. This was the therapy with the largest number of patients included in the study. **Error! Reference source not found.**

The majority of patients were female (72%), white (71%) and had no other past medical history (52%). The median age of MS onset was 30 years and the most common phenotype at presentation was a spinal cord presentation (35%). Only 33% of patients had a lumbar puncture, with the majority (93%) showing evidence of unmatched OCBs in the CSF as compared to the serum. Around 50% of patients were DMT naïve and 37% had one previous DMT prior to starting DMF. The median age at starting DMF was 40 years and the median disease duration prior to commencing DMF was 7 years. The median number of relapses occurring in the 12 months prior to starting DMF was 1 relapse and the median number of new MS lesions on baseline MRI brain was 0 (range: 0-10). The median EDSS score at baseline was 2.0. Clinical and baseline demographic data are shown in Table 16: DMF: Demographics and baseline clinical details.

Table 16: DMF: Demographics and baseline clinical details.			
Sex		**Age at DMT initiation (years)**	
Female (%)	**483 (72)**	Median	**40**
Male	187	Mean	40.5
		Min	16
		Max	82
Ethnic background		**No. of previous DMTs**	
White (%)	**478 (71)**	0	**335 (50)**
Black	29	1	250 (37)
Asian	65	2	62
Mixed	24	3	15
Other	22	4	8
Presence of other comorbidities		**Disease duration (years)**	
		Median	**7**
Yes (%)	324 (48)	Mean	8.6
No	346	Min	0
		Max	40
Age at MS onset (years)		**No. Relapses in previous 12 months**	
Median	**30**	0	236
Mean	32	1	308
Min	10	2	104
Max	73	3	21
Phenotype at MS onset		**No. new MRI lesions at DMT initiation**	
Spinal cord (SC) (%)	**236 (35)**	Median	**0**
Optic neuritis (ON)	155	Mean	0.8
Supratentorial (ST)	111	Min	0
Infratentorial (IT)	124	Max	10
Other	9		
Genetics (%)	106/670 (16)	**EDSS at DMT initiation**	
0 = no alleles	9	Median	**2.0**
1 = one allele	13	Mean	2.6
2 = two alleles	0	Min	0.0
Pending phenotype	84	Max	7.5
OCBs (%)	221/670 (33)	**Duration on DMF**	
Yes (%)	206/221 (93**)**	Average years (range)	2.5 (0-8)
No (%)	15/221 (7)		

Only 106 (16%) patients had undergone serum genetic analysis for HLA-DRB1*15 and the results of only 22 patients have been phenotyped to date. Further analysis of this data has been postponed due to a lack of laboratory analysis which was affected by the COVID-19 pandemic. Other missing data are summarised in Table 17: Missing data in the dimethyl fumarate cohort. Some variables are not included in the table as no missing data was recorded.

Table 17: Missing data in the dimethyl fumarate cohort

	Baseline (%)	**12 months (%)**	**24 months (%)**
Ethnicity	52/670 (7)	N/A	N/A
Phenotype at MS onset	36/670 (5)	N/A	N/A
OCBs	449/670 (67)	N/A	N/A
Genetics	564/670 (84)	N/A	N/A
Relapses	0	60 (9)	150 (22)
EDSS	0	61 (9)	152 (22)
MRI	15 (2)	139 (20)	227 (40)

NEDA was calculated at 12 and 24 months for the DMF cohort and the results are summarised in Table 18: DMF: NEDA at 12 months and 24 months. Out of all the patients who initiated DMF, 76% achieved NEDA at 12 months and 58% achieved NEDA at 24 months. Approximately, 22% (148) and 29% (194) of patients stopped therapy by the end of year 1 and 2 respectively. The causes of patients stopping DMF in the first 2 years are summarised in **Error! Reference source not found.**.

Table 18: DMF: NEDA at 12 months and 24 months.	
NEDA at 12 months	
Yes (%)	510/670 (76)
No (%)	160/670 (24)
NEDA at 24 months	
Yes (%)	387/670 (58)
No (%)	283/670 (42)

Table 19: Causes of stopping DMF				
Causes	**Patients stopping DMT within 6 months**	**Patients stopping DMT within 12 months**	**Patients stopping DMT within 18 months**	**Patients stopping DMT within 24 months**
Side effects	57	93	94	102
Disease activity	2	26	29	39
Lymphopenia	0	11	11	21
Family planning	0	7	7	17
Lost to follow up	1	11	11	15
Total	60	148	152	194

3.4.1.2 Fingolimod

A total of 336 patients were initiated on fingolimod, a second line therapy, between 2011 and February 2020. Clinical and baseline demographic data are shown in Table 20: Fingolimod: Demographics and baseline clinical details.

The majority of patients were female (75%), white (75%) and had a history of other comorbidities (60%). The median age of MS onset was 28 years and the most common clinical phenotype at MS onset was a spinal cord presentation (32%). Similar to DMF, only 36% of patients had a lumbar puncture and the majority (93%) showed evidence of unmatched OCBs in the CSF as compared to the serum. More than half of patients (55%) had trialled one DMT prior to commencing fingolimod and the majority of the remaining patients had trialled more than one DMT. This is to be expected as fingolimod is a second line therapy in the UK. The median age at starting fingolimod was 39 years and the median disease duration prior to commencing fingolimod was 9 years. The median number of relapses occurred in the previous 12 months of starting therapy was 1 relapse and the median number of new MS lesions

on baseline MRI brain was 0 (range: 0-13). The median EDSS score at baseline was 2.5.

Table 20: Fingolimod: Demographics and baseline clinical details.			
Sex		**Age at DMT initiation (years)**	
Female (%)	**252 (75)**	Median	39
Male	84	Mean	39.2
		Min	16
		Max	69
Ethnic background		**No. of previous DMTs**	
White (%)	**243 (75)**	0	12
Black	16	1	**185 (55)**
Asian	35	2	106
Mixed	15	3	30
Other	12	4	2
Presence of other comorbidities		**Disease duration (years)**	
		Median	9
Yes (%)	200 (60)	Mean	10.6
No	134	Min	0
		Max	48
Age at MS onset (years)		**No. Relapses in previous 12 months**	
Median	28		
Mean	28.6	0	86
Min	8	1	**160**
Max	56	2	74
		3	10
		4	3
Phenotype at MS onset		**No. new MRI lesions at DMT initiation**	
SC = spinal cord (%)	106 (32)		
ON = optic neuritis	73	0	**438**
ST = supratentorial	48	1	73
IT = infratentorial	76	2	55
Other	2	3	49
		4	20
		5 or more	20
Genetics (%)	73/336 (22)	**EDSS (T0)**	
0 = no alleles	19	Median	**2.5**
1 = one allele	8	Mean	3.5
2 = two alleles	4	Min	0.0
Pending phenotype	42	Max	7.5
OCBs (%)	120/336 (36)	**Duration on fingolimod**	
Yes (%)	112/120 (93)	Average years (range)	4.4 (0-10)
No (%)	8/120 (7)		

Only 73 (22%) patients had undergone serum genetic analysis for HLA-DRB1*15 and only 31 patients have been phenotyped to date. Further analysis of this data has therefore been postponed due to the lack of samples being processed in the laboratory since the impact of the COVID-19 pandemic. Other missing data are summarised in Table 21: Missing data in the fingolimod cohort. Some variables are not included in the table as no missing data was recorded.

Table 21: Missing data in the fingolimod cohort			
	Baseline (%)	**12 months (%)**	**24 months (%)**
Ethnicity	5/336 (2)	N/A	N/A
Phenotype at MS onset	32/336 (9)	N/A	N/A
OCBs	216/336 (64)	N/A	N/A
Genetics	263/336 (78)	N/A	N/A
Relapses	3	19 (5)	51 (15)
EDSS	0	24 (7)	53 (15)
MRI	27 (8)	98 (29)	113 (33)

The results for NEDA were calculated for this cohort at 12 and 24 months, the outcome is summarised in Table 22: Fingolimod: NEDA at 12 months and 24 months. For all patients who commenced fingolimod, 78% achieved NEDA at 12 months and 61% achieved NEDA at 24 months. Approximately, 15% (49) and 25% (84) of patients stopped therapy by the end of year 1 and 2, respectively. The causes of patients stopping fingolimod in the first 2 years are summarised in Table 23: Causes of stopping Fingolimod

Table 22: Fingolimod: NEDA at 12 months and 24 months.	
NEDA at 12 months	
Yes (%)	262/336 (78)
No	74/336 (22)
NEDA at 24 months	
Yes (%)	206/336 (61)
No	130/336 (39)

Table 23: Causes of stopping Fingolimod				
Causes	**Patients stopping DMT within 6 months**	**Patients stopping DMT within 12 months**	**Patients stopping DMT within 18 months**	**Patients stopping DMT within 24 months**
Side effects				
- **Deranged LFTs**	3	7	7	9
- **Cardiac causes**	7	8	8	10
- **Macular causes**	2	4	4	4
- **Headaches**	2	3	3	3
- **Infections**	4	6	6	6
- **Other**			1	5
Disease activity	0	13	14	28
Lymphopenia	0	1	1	3
Family planning	0	2	2	6
Lost to follow up	1	5	5	10
Total	19	49	51	84

3.4.1.3 Natalizumab

Natalizumab is a highly effective DMT. Between 2007 and October 2019, 177 patients were started on natalizumab. The baseline demographic and clinical data can be seen in Table 24: Natalizumab: Demographics and baseline clinical details.

The majority of patients in this cohort were female (64%), white (67%) and had other comorbidities (72%). The median age of MS onset was 27 years and there were three clinical phenotypes at MS onset with similar prevalence within the cohort; spinal cord (23%), optic neuritis (22%) and infratentorial (21%). Only 35% of patients had a lumbar puncture with OCBs results available and the majority of those (95%) showed evidence of unmatched OCBs in the CSF compared to the serum. As expected the majority of patients had tried at least one DMT prior to starting natalizumab, however, interestingly 25 (14%) patients were treatment naïve. The median age at starting natalizumab was 36 years and the median disease duration prior to commencing natalizumab was 7 years. The median number of relapses occurring in the 12 months prior to starting natalizumab was 2 relapses and the median number of new MS lesion on baseline MRI brain was 1 (range: 0-8). The median EDSS score at baseline was 6.0. These features demonstrate a cohort with higher disability and more active MS compared to the previous cohorts. This is not unexpected considering natalizumab is typically prescribed for highly active RRMS and for those who have failed first- and second-line therapies.

Table 24: Natalizumab: Demographics and baseline clinical details.			
Sex		**Age at DMT initiation (year)**	
Female (%)	**113 (64)**	Median	36
Male	64	Mean	36
		Min	15
		Max	68
Ethnic background		**No. of previous DMTs**	
White (%)	118 (67)	0	25
Black	20	1	**91 (51)**
Asian	16	2	38
Mixed	6	3	19
Other	11	4	3
Presence of other comorbidities		**Disease duration (years)**	
		Median	**7**
Yes (%)	128 (72)	Mean	8.4
No	48	Min	0
		Max	41
Age at MS onset (years)		**No. Relapses 12m prior to DMT initiation**	
Median	27		
Mean	27.7	Median	**2**
Min	10	Mean	1.9
Max	65	Min	0
		Max	6
Phenotype at MS onset		**No. new MRI lesions at DMT initiation**	
Spinal cord (%)	41 (23)		
Optic neuritis	38 (22)	Median	**1**
Supratentorial	19 (11)	Mean	1.7
Infratentorial	37 (21)	Min	0
Other	7	Max	8
Genetics (%)	106/177 (60)	**EDSS at DMT initiation**	
0 = no alleles	28	Median	**6.0**
1 = one allele	1	Mean	4.7
2 = two alleles	7	Min	1.0
Pending phenotype	57	Max	8.0
OCBs	61/177	**Duration on natalizumab**	
Yes (%)	58/61 (95)	Average years (range)	7 (0.5-15)
No	3/61 (5)		

Only 106/176 (60%) patients had undergone serum genetic analysis for HLA-DRB1*15 to date and only 36 patients were phenotyped. Further analysis of this data has therefore been postponed due to the lack of sample processing and phenotyping since the impact of the COVID-19 pandemic. Other missing data is summarised in

Table 25: Missing data in the natalizumab cohort. Some variables are not included in the table as no missing data was recorded.

Table 25: Missing data in the natalizumab cohort			
	Baseline (%)	**12 months (%)**	**24 months (%)**
Ethnicity	5/177 (3)	N/A	N/A
Phenotype at MS onset	14/177 (8)	N/A	N/A
OCBs	116/177 (65)	N/A	N/A
Genetics	71/177 (40)	N/A	N/A
Relapses	0	0	8 (4)
EDSS	0	0	8 (4)
MRI	15 (8)	9 (5)	9 (5)

Having calculated NEDA for the natalizumab cohort at 12 and 24 months after commencing therapy, the results are summarised in Table 26: Natalizumab: NEDA at 12 months and 24 months. Out of all the patients who received natalizumab, 81% (144) achieved NEDA at 12 months and 69% (123) achieved NEDA at 24 months.

Table 26: Natalizumab: NEDA at 12 months and 24 months.	
NEDA at 12 months	
Yes (%)	144/177 (81) – total cohort
No	33/177 (19)
NEDA at 24 months	
Yes (%)	123/177 (69) – total cohort
No	54/177 (31)

There was a variety of reasons why patients stopped therapy in the first year, however it was the expected increased risk of PML that influenced the majority of patients who discontinued therapy at 2 years. The causes of patients stopping natalizumab therapy by the end of year 1 and 2 are summarised in Table 27: Causes of stopping natalizumab

Table 27: Causes of stopping natalizumab				
Causes	**Patients stopping DMT within 6 months**	**Patients stopping DMT within 12 months**	**Patients stopping DMT within 18 months**	**Patients stopping DMT within 24 months**
Side effects	0	2	2	2
PML risk	0	1	1	7
Disease activity	0	0	0	2
Compliance	0	2	2	2
Family planning	0	2	2	2
Lost to follow up	0	1	1	1
Total	0	8	8	16

3.4.1.4 Glatiramer acetate

Within the study there were 547 patients starting glatiramer acetate therapy between 2002 and March 2021. Glatiramer acetate is a first line moderately effective DMT.

The majority of patients were female (74%), white (78%) and had a history of other comorbidities (63%). The median age of MS onset was 30 years and the most common clinical phenotype at MS onset was a spinal cord presentation (38%). Similarly, only 35% of patients had a lumbar puncture with OCBs results available and the majority of these (92%) showed evidence of unmatched OCBs in the CSF as compared to the serum. The majority of patients (63%) were treatment naïve prior to starting glatiramer acetate (GA), which is expected considering GA is a first line therapy which has been approved in the UK for over two decades. The median age at starting GA was 37 years and the median disease duration prior to commencing GA was 5 years. Three quarters of patients had no or one relapse in the previous 12 months. The median number of new MS lesions on baseline MRI brain was 0. The median EDSS score at baseline was 2.0. These results identify that this cohort demonstrates low MS activity and disability, which is expected considering GA is only moderately effective. Clinical and baseline demographic data are shown in Table 28: Glatiramer acetate: Demographics and baseline clinical details.

Table 28: Glatiramer acetate: Demographics and baseline clinical details.			
Sex		**Age at DMT initiation (years)**	
Female (%)	**407 (74)**	Median	37
Male	140	Mean	38.7
		Min	16
		Max	72
Ethnic background		**No. of previous DMTs**	
White (%)	**425 (78)**	0	**342 (63)**
Black	33	1	150
Asian	43	2	48
Mixed	17	3	6
Other	15	4	1
Presence of other comorbidities		**Disease duration (years)**	
		Median	5
Yes (%)	347 (63)	Mean	7.3
No	186	Min	0
		Max	36
Age at MS onset (years)		**No. Relapses in previous 12 months**	
Median	30	0	149
Mean	31.4	1	251
Min	8	2	116
Max	67	3	25
		4	3
Phenotype at MS onset		**No. new MRI lesions at DMT initiation**	
SC = spinal cord (%)	210 (38)	0	**187**
ON = optic neuritis	128	1	41
ST = supratentorial	82	2	40
IT = infratentorial	98	3	31
		4	28
		5-9	36
Genetics (%)	85/547 (16)	**EDSS (T0)**	
0 = no alleles	7	Median	**2.0**
1 = one allele	15	Mean	2.5
2 = two alleles	0	Min	0
Pending phenotype	63	Max	7.0
OCBs (%)	191/547 (35)	**Duration on GA**	
Yes (%)	175/191 (92)	Average years (range)	4.7 (0-20)
No (%)	16/191 (8)		

Only 85 (16%) patients had undergone serum genetic analysis for HLA-DRB1*15 to date and only 22 patients were phenotyped. Further analysis of this data has been postponed due to the lack of samples being processed in the laboratory due to the

COVID-19 pandemic impact. Other missing data are summarised in Table 29: Missing data in the glatiramer acetate cohort. Some variables are not included in the table as no missing data was recorded. It is also clear that there is a significant amount of MRI scans missing in this cohort. This is mainly due to the fact that previously patients on this therapy were not scanned regularly, unless symptomatic, due to the high safety profile and the lack of DMTs available for radiological activity in the absence of clinical relapses or deterioration. The guidance has changed in the last few years, as all MS patients on DMTs are routinely having annual MRI scans.

Table 29: Missing data in the glatiramer acetate cohort			
	Baseline (%)	**12 months (%)**	**24 months (%)**
Ethnicity	14/547 (2)	N/A	N/A
Phenotype at MS onset	29/547 (5)	N/A	N/A
OCBs	356/547 (65)	N/A	N/A
Genetics	462/547 (84)	N/A	N/A
Relapses	0	40/547 (7)	111/532 (20)
EDSS	0	44/547 (8)	117/532 (22)
MRI	184/547 (33)	367/547 (67)	370/532 (69)

Having calculated NEDA for this cohort at 12 and 24 months, the results are summarised in Table 30: Glatiramer acetate: NEDA at 12 months and 24 months. Out of all the patients who commenced GA (547), 65% achieved NEDA at 12 months. Out of all the patients who initiated therapy before September 2020 and were eligible for NEDA assessment at 24 months (532), 248 patients (46%) achieved NEDA.

Approximately, 22% (119) of patients stopped GA by the end of the first year and 34% (180) of patients stopped GA by the end of the second year. The causes of patients stopping GA in the first 2 years are summarised in Table 31: Causes of stopping glatiramer acetate. It was noted that 20 patients discontinued therapy in the first two years for family planning. This number is expected to be lower in recent years considering recent data and UK consensus developed through the association of

British neurologists in 2019 reporting that GA is safe during conception, pregnancy and breastfeeding. (213)

Table 30: Glatiramer acetate: NEDA at 12 months and 24 months.	
NEDA at 12 months	
Yes (%)	358/547 (65)
No	189/547 (35)
NEDA at 24 months	
Yes (%)	248/532 (46.6)
No	284/532 (53.4)

Table 31: Causes of stopping glatiramer acetate				
Causes	**Patients stopping DMT within 6 months**	**Patients stopping DMT within 12 months**	**Patients stopping DMT within 18 months**	**Patients stopping DMT within 24 months**
Side effects	31	63	67	76
Disease activity	0	34	34	55
Compliance	0	3	3	8
Family planning	1	10	13	20
Lost to follow up	2	9	9	21
Total	34	119	126	180

3.4.1.5 Ocrelizumab

Ocrelizumab is a highly effective DMT. Between July 2019 and February 2021, 256 patients commenced ocrelizumab in the PITMS prospective study.

The majority of patients were female (63%), white (75%) and had a history of other comorbidities (56%). The median age of MS onset was 30 years and the most common clinical phenotype at MS onset was a spinal cord presentation (40%). The median age at starting ocrelizumab was 38 years and the median disease duration prior to commencing treatment was 5 years. Interestingly, the majority of patients (54%) were treatment naïve prior to starting ocrelizumab, however, 70% had relapses in the previous 12 months. Just under half of the patients had no new MS lesions on baseline MRI brain, whereas the other half had a range of 1-13 new MS lesions. The median EDSS score at baseline was 2.0. These results show a variety of patients with low and high MS activity and disability, this is likely due to the fact that ocrelizumab has a much wider eligibility criteria due to its high efficacy and relatively low risk profile. Clinical and baseline demographic data are shown in Table 32: Ocrelizumab: Demographics and baseline clinical details.

Having calculated NEDA for this cohort at 12 and 24 months, out of all the prospective PITMS patients who commenced ocrelizumab (256), 92% achieved NEDA at 12 months. Ocrelizumab appears to be well tolerated with only one patient stopping therapy during cycle one due to infusion-related-reactions. For NEDA assessment at 24 months, we only included 120 patients who initiated therapy up to August 2020; 87/120 (72.5%) achieved NEDA. The results are summarised in Table 33: Ocrelizumab: NEDA at 12 months and 24 months.

Table 32: Ocrelizumab: Demographics and baseline clinical details			
Sex		**Age at DMT initiation (years)**	
Female (%)	**160 (63)**	Median	38
Male	96	Mean	38.8
		Min	18
		Max	68
Ethnic background		**No. of previous DMTs**	
White (%)	**193 (75)**	0	**138 (54)**
Black	16	1	54
Asian	27	2	40
Mixed	14	3	15
Other	6	4 or more	9
Presence of other comorbidities		**No. Relapses in previous 12 months**	
Yes (%)	144 (56)	0	75
No	112	1	120
		2	53
		3	8
Age at MS onset (years)		**No. new MRI lesions at DMT initiation**	
Median	30	0	**119 (46)**
Mean	31.7	1	39
Min	11	2	27
Max	65	3	13
		4	16
		5-9	19
		10 or more	3
Phenotype at MS onset		**EDSS (T0)**	
SC = spinal cord (%)	**102 (40)**	Median	2.0
ON = optic neuritis	61	Mean	2.7
ST = supratentorial	23	Min	0
IT = infratentorial	60	Max	8.0
Disease duration (years)		**Duration on ocrelizumab**	
Median	5	Average months (range)	26.7(0-31)
Mean	7.1		
Min	0		
Max	30		

Table 33: Ocrelizumab: NEDA at 12 months and 24 months	
NEDA at 12 months	
Yes (%)	235/256 (92) – total cohort
No	21/256 (8)
NEDA at 24 months	
Yes (%)	87/120 (72.5%) – eligible cohort
No	33/120 (27.5%)

In total, seven (3%) patients stopped therapy by the end of year one. Out of all the patients (120) who were eligible for NEDA assessment at 2 years, five patients stopped therapy within 18 months; due to infusion-related-reactions (1), undergoing cancer investigations (1), elected to have stem cell transplant (1), family planning (1), and lost to follow up (1). None of the patients stopped therapy due to disease activity. Causes of all patients in the ocrelizumab cohort stopping within 24 months are summarised in Table 34: Causes of stopping ocrelizumab.

Table 34: Causes of stopping ocrelizumab	
< 6 months	Side effects = 1 (infusion-related-reaction on dose 1a)
0-12 months	Side effects = 1, family planning = 3, lost to follow up = 1, others = 2 [Total = 7]
0-24 months	Side effects = 1, family planning = 3, lost to follow up = 2, others = 3 [Total = 9]

Missing data are summarised in Table 35: Missing data in the ocrelizumab cohort. One of the major reasons to the significant number of missing scans was the COVID pandemic and many patients could not attend for their MRI appointments.

Table 35: Missing data in the ocrelizumab cohort			
	Baseline (%)	**12 months (%)**	**24 months (%)**
Ethnicity	0	N/A	N/A
Phenotype at MS onset	10/256 (4)	N/A	N/A
OCBs	N/A	N/A	N/A
Genetics	N/A	N/A	N/A
Relapses	0	1	5/107 (5)
EDSS	0	4/256 (2)	6/107 (6)
MRI	20/256 (8)	65/256 (25)	35/107 (32)

3.4.2 Predicting NEDA using combined multiple logistic regression

3.4.2.1 Dimethyl fumarate (DMF) Cohort:

In modelling NEDA at 12 months, we used most baseline variables and excluded those with missing data of more than 10% including genetics and OCBs. In our prediction model we only included patients who received therapy for ≥6 months; 610/670 patients (91%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years.

Logistic regression (LR) coefficient estimates give the change in the log-odds of the outcome for a one unit increase in the predictor variable, assuming all other variables are constant at a certain value. This is summarised in Table 37: Dimethyl fumarate predictors for NEDA at 12 months. Notably, the number of relapses in the previous 12 months, the number of new MS lesions on the baseline MRI brain and the EDSS score at baseline were significantly associated with achieving NEDA at 12 months. Importantly, these coefficients are negative and can therefore be interpreted as any increase in these variables will reduce the chances of achieving NEDA. A supratentorial (ST) clinical presentation at MS onset showed a statistically significant strongly positive coefficient estimate in achieving NEDA compared to an infratentorial (IT) presentation.

The model findings for NEDA at 24 months is shown in Table 38: DMF predictors for NEDA at 24 months. In our prediction model we only included patients who received therapy for ≥18 months; 547/670 patients (82%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years. Compared to predictors at one year, relapses in the previous 12 months of initiating DMF, number of new MS lesions at baseline, and EDSS score at baseline continued to be statistically significant predictor of achieving NEDA at 24 months. This shows the importance of these predictors for achieving NEDA at short and medium terms. It is also interesting to note that age at DMF initiation, the number of previous DMTs and the presence of other comorbidities were also statistically significant and are therefore important NEDA predictors to be considered in the longer term.

Overall, these findings show that in our cohort DMF has a greater chance of achieving NEDA in patients with certain baseline demographics; low number of relapses in the preceding 12 months, low number of new MS lesions on the baseline MRI brain scan, and low EDSS at baseline. It is also important to consider age, comorbidities and being treatment naïve, at least in the medium term, for achieving NEDA.

Table 36: Number of patients assessed at 1 and 2 years

	Glatiramer Acetate	**Dimethyl fumarate**	**Fingolimod**	**Natalizumab**	**Ocrelizumab**	**Total number**
Number of patients initiated DMT	547	670	336	177	256	1986
Eligible for assessment at 1 year (%)	514 (94)	610 (91)	317 (94)	177 (100)	255 (99)	1873
Number of patients at 2 years	532	670	336	177	120	1835
Eligible for assessment at 2 years (%)	458 (86)	547 (82)	299 (89)	169 (95)	115 (96)	1588

Table 37: Dimethyl fumarate predictors for NEDA at 12 months

	Coefficient (logOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	1.65	0.65	2.56	0.01
Age at DMF initiation	0.02	0.01	1.27	0.21
Sex: Male	0.11	0.26	0.44	0.66
Ethnicity: Black	0.11	0.57	0.19	0.85
Ethnicity: Mixed	1.21	0.84	1.45	0.15
Ethnicity: Other	0.17	0.66	0.26	0.79
Ethnicity: White	0.28	0.34	0.83	0.41
Phenotype: Optic neuritis	-0.05	0.33	-0.14	0.89
Phenotype: Spinal cord	-0.02	0.32	-0.08	0.94
Phenotype: Supratentorial	0.94	0.46	2.05	**0.04***
Phenotype: other	-1.31	0.80	-1.64	0.10
Presence of other comorbidities	-0.32	0.24	-1.34	0.18
No. of previous DMTs	-0.12	0.15	-0.79	0.43
Disease duration	0.04	0.02	1.80	0.07
No. of relapses in previous 12m	-0.44	0.15	-3.00	**0.003****
No. of new MRI lesions at DMT initiation	-0.16	0.073	-2.24	**0.03***
EDSS at DMT initiation	-0.20	0.07	-2.77	**0.005****

Table 38: DMF predictors for NEDA at 24 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	0.711	0.57	1.25	0.21
Age DMT initiation	0.04	0.01	3.25	**0.001****
Sex: Male	-0.20	0.22	-0.87	0.38
Ethnicity: Black	0.14	0.51	0.28	0.78
Ethnicity: Mixed	0.50	0.63	0.80	0.43
Ethnicity: Other	-0.32	0.53	-0.60	0.55
Ethnicity: White	0.17	0.31	0.56	0.58
Phenotype: Optic neuritis	-0.25	0.30	-0.84	0.40
Phenotype: Spinal cord	-0.30	0.28	-1.07	0.28
Phenotype: Supratentorial	0.28	0.35	0.81	0.42
Phenotype: Other	-1.35	0.78	-1.75	0.08
Presence of other comorbidities	-0.51	0.20	-2.50	**0.01***
No. of previous DMTs	-0.32	0.13	-2.37	**0.02***
Disease duration	0.02	0.02	0.97	0.33
No. of relapses in previous 12 months	-0.47	0.13	-3.66	**0.0003*****
No. of new MRI lesions at DMT initiation	-0.13	0.07	-2.02	**0.04***
EDSS at DMT initiation	-0.17	0.07	-2.65	**0.01***

3.4.2.2 Fingolimod Cohort:

The outcome of modelling NEDA at 12 months for Fingolimod using most variables and excluding those with missing data of more than 10%, including genetics and OCBs, is summarised in Table 39: Fingolimod predictors for NEDA at 12 months. In our prediction model we only included patients who received therapy for ≥6 months; 317/336 patients (94%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years. Notably, although none of the coefficients in our model reached statistical significance, the coefficients for age at fingolimod initiation only just missed the statistically significant cut off. Again, although it did not reach statistical significance, it is interesting to see that the coefficient of the number of new MRI lesions at baseline is now positive. This can be interpreted that NEDA was more likely to be achieved in those starting fingolimod with radiologically active MS. However, this was not replicated at 24 months.

In our prediction model for NEDA at 24 months, we only included patients who received therapy for ≥18 months; 299/336 patients (89%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years. The outcome of modelling NEDA at 24 months is summarised in Table 40: Fingolimod predictors for NEDA at 24 months. It highlighted that the number of relapses in the previous 12 months of starting fingolimod is statistically significant in predicting NEDA at 24 months. Although other variables did not reach significance, most variables had similar coefficient directions (positive/negative).

Table 39: Fingolimod predictors for NEDA at 12 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	2.95	0.87	3.37	0.001
Age at DMT initiation	-0.03	0.02	-1.66	**0.09**
Sex: Male	-0.03	0.37	-0.08	0.94
Ethnicity: Black	1.89	1.16	1.63	0.10
Ethnicity: Mixed	-1.15	0.75	-1.52	0.13
Ethnicity: Other	1.02	1.14	0.90	0.37
Ethnicity: White	0.73	0.46	1.57	0.12
Phenotype: Optic neuritis	0.21	0.44	0.48	0.63
Phenotype: Spinal cord	0.38	0.41	0.92	0.35
Phenotype: Supratentorial	0.85	0.58	1.46	0.14
Presence of other comorbidities	-0.46	0.34	-1.33	0.18
No. of previous DMTs	-0.20	0.22	-0.90	0.37
Disease duration	0.03	0.03	1.22	0.22
No. of relapses in previous 12 months	-0.36	0.20	-1.81	**0.07**
No. of new MRI lesions at DMT initiation	0.09	0.10	0.91	0.37
EDSS at DMT initiation	-0.12	0.09	-1.26	0.21

Table 40: Fingolimod predictors for NEDA at 24 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	1.37	0.73	1.87	0.06
Age at DMT initiation	0.02	0.02	1.10	0.27
Sex: Male	-0.01	0.31	-0.03	0.98
Ethnicity: Black	0.26	0.72	0.37	0.71
Ethnicity: Mixed	-0.95	0.74	-0.129	0.20
Ethnicity: Other	0.64	0.88	0.73	0.46
Ethnicity: White	0.18	0.42	0.43	0.67
Phenotype: Optic neuritis	-0.38	0.37	-1.02	0.31
Phenotype: Spinal cord	-0.22	0.36	-0.61	0.54
Phenotype: Supratentorial	0.28	0.47	0.59	0.55
Presence of other comorbidities	-0.25	0.28	-0.89	0.37
No. of previous DMTs	-0.16	0.19	-0.86	0.39
Disease duration	-0.003	0.02	-0.15	0.87
No. of relapses in previous 12 months	-0.32	0.16	-1.95	**0.05***
No. of new MRI lesions at DMT initiation	-0.02	0.09	-0.20	0.84
EDSS at DMT initiation	-0.13	0.08	-1.62	0.11

3.4.2.3 Natalizumab Cohort:

The outcome of modelling NEDA at 12 months for the Natalizumab cohort using most variables and excluding those with missing data of more than 10%, including genetics and OCBs, is summarised in Table 41: Natalizumab predictors for NEDA at 12 months. In our prediction model we included all 177 patients (100%). Our model showed that number of relapses in the previous 12 months and EDSS score reached statistical significance. Interestingly, the coefficient direction of the number of relapses in the previous 12 months was positive. This possibly reflects the reality that natalizumab is usually given to those with more active MS (with higher number of relapses), but again more likely to achieve NEDA in those with low disability (lower EDSS score). This was also replicated at 24 months.

In our prediction model for NEDA at 24 months, we only included patients who received therapy for ≥18 months; 169/177 patients (95%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years. The outcome of modelling NEDA at 24 months is summarised in Table 42: Natalizumab predictors for NEDA at 24 months. This showed that age at starting natalizumab, MS duration prior to starting therapy and EDSS at therapy initiation were statistically significant predictors of reaching NEDA at 24 months in this cohort.

Table 41: Natalizumab predictors for NEDA at 12 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	2.85	1.54	1.85	0.07
Age at DMT initiation	-0.03	0.02	-1.31	0.20
Sex: Male	-0.04	0.45	-0.10	0.93
Ethnicity: Black	0.43	0.98	0.44	0.66
Ethnicity: Mixed	15.91	1497	0.01	0.99
Ethnicity: Other	0.44	1.16	0.38	0.70
Ethnicity: White	0.21	0.80	0.27	0.80
Phenotype: Optic neuritis	0.50	0.60	0.81	0.42
Phenotype: Spinal cord	0.84	0.63	1.33	0.18
Phenotype: Supratentorial	-0.16	0.65	-0.25	0.81
Presence of other comorbidities	-0.20	0.51	-0.40	0.70
No. of previous DMTs	-0.20	0.25	-0.75	0.45
Disease duration	0.04	0.04	1.04	0.30
No. of relapses in previous 12 months	0.48	0.25	1.92	**0.05***
No. of new MRI lesions at DMT initiation	0.17	0.12	1.48	0.14
EDSS at DMT initiation	-0.34	0.13	-2.63	**0.01***

Table 42: Natalizumab predictors for NEDA at 24 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	4.02	1.50	2.70	0.01
Age at DMT initiation	-0.06	0.02	-2.65	**0.01***
Sex: Male	0.31	0.42	0.73	0.46
Ethnicity: Black	-0.03	1.12	-0.03	0.98
Ethnicity: Mixed	-0.60	1.53	-0.40	0.70
Ethnicity: Other	-1.34	1.16	-1.15	0.25
Ethnicity: White	-0.87	0.93	-0.93	0.35
Phenotype: Optic neuritis	0.56	0.53	1.06	0.30
Phenotype: Spinal cord	1.04	0.56	1.86	0.06
Phenotype: Supratentorial	0.33	0.62	0.53	0.60
Presence of other comorbidities	-0.47	0.47	-1.01	0.31
No. of previous DMTs	-0.35	0.23	-1.52	0.13
Disease duration	0.08	0.04	2.17	**0.03***
No. of relapses in previous 12 months	0.30	0.21	1.40	0.16
No. of new MRI lesions at DMT initiation	0.10	0.10	1.03	0.30
EDSS at DMT initiation	-0.23	0.11	-2.13	**0.03***

3.4.2.4 Glatiramer acetate cohort:

The outcome of modelling NEDA at 12 months for the glatiramer acetate (GA) cohort using most variables and excluding those with missing data of more than 50%, including genetics and OCBs, is summarised in Table 43: Glatiramer acetate predictors for NEDA at 12 months. In our prediction models we only included patients who received therapy for ≥6 months; 513/547 patients (94%), and ≥18 months; 458/532 patients (86%) for predicting NEDA at 12 and 24 months respectively. This is summarised in Table 36: Number of patients assessed at 1 and 2 years.

Our model showed that age at GA initiation, all clinical phenotypes compared to infratentorial (IT) presentation, number of previous DMTs used, number of relapses in the previous 12 months, number of new MRI lesions and EDSS score at therapy initiation were all statistically significant predictors for achieving NEDA at 12 months in this GA group. Although not all reached statistical significance, those predictors had similar trends at predicting NEDA at 24 months. Once again, in this first-line therapy cohort, the probability of achieving NEDA was more likely to occur in those with low MS activity; treatment naïve, lower number of relapses, lower number of new MRI lesions and lower EDSS score.

Table 43: Glatiramer acetate predictors for NEDA at 12 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	0.17	0.64	0.26	0.80
Age at DMT initiation	0.04	0.01	2.55	**0.01***
Sex: Male	0.43	0.25	1.74	**0.08**
Ethnicity: Black	-0.80	0.53	-1.51	0.13
Ethnicity: Mixed	-0.80	0.63	-1.26	0.21
Ethnicity: Other	0.33	0.68	0.50	0.63
Ethnicity: White	0.17	0.36	0.47	0.64
Phenotype: Optic neuritis	0.66	0.32	2.10	**0.04***
Phenotype: Spinal cord	0.65	0.30	2.28	**0.02***
Phenotype: Supratentorial	1.30	0.38	3.41	**0.0006****
Presence of other comorbidities	-0.12	0.23	-0.53	0.60
No. of previous DMTs	-0.32	0.16	-1.97	**0.05***
Disease duration	0.002	0.02	0.11	0.91
No. of relapses in previous 12 months	-0.27	0.13	-2.14	**0.03***
No. of new MRI lesions at DMT initiation	-0.15	0.05	-2.77	**0.005****
EDSS at DMT initiation	-0.27	0.07	-4.07	**0.0001*****

Table 44: Glatiramer acetate predictors for NEDA at 24 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	-0.22	0.66	-0.33	0.74
Age at DMT initiation	0.03	0.01	2.16	**0.03***
Sex: Male	0.34	0.23	1.47	0.14
Ethnicity: Black	-0.65	0.57	-1.12	0.26
Ethnicity: Mixed	-0.65	0.68	-0.95	0.34
Ethnicity: Other	0.16	0.90	0.18	0.86
Ethnicity: White	-0.37	0.40	-0.91	0.36
Phenotype: Optic neuritis	0.44	0.32	1.40	0.16
Phenotype: Spinal cord	0.62	0.29	2.13	**0.03***
Phenotype: Supratentorial	0.88	0.35	2.54	**0.01***
Presence of other comorbidities	0.17	0.22	0.74	0.46
No. of previous DMTs	-0.12	0.17	-0.70	0.50
Disease duration	-0.02	0.02	-0.88	0.38
No. of relapses in previous 12 months	-0.11	0.13	-0.83	0.41
No. of new MRI lesions at DMT initiation	-0.13	0.05	-2.43	**0.01***
EDSS at DMT initiation	-0.21	0.07	-3.11	**0.002****

3.4.2.5 Ocrelizumab cohort:

The outcome of modelling NEDA at 12 months for the ocrelizumab cohort using most variables is summarised in Table 45: Ocrelizumab predictors for NEDA at 12 months. In our prediction model we only included patients who received therapy for ≥6 months 255/256 patients. Our model showed that age at ocrelizumab initiation, number of previously used DMTs, disease duration and number of new MRI lesions were statistically significant predictors of NEDA at 12 months. This can be interpreted that NEDA in this cohort was most likely to be achieved in those who were older, treatment naïve, initiated treatment early and had no or low MRI brain activity at baseline. It also showed those with high relapse rate were more likely to achieve NEDA, however this variable only just missed the cut off for statistical significance.

The outcome of modelling NEDA at 24 months is summarised in Table 46: Ocrelizumab predictors for NEDA at 24 months. In our prediction model we only included patients who received therapy for ≥18 months 115/120 patients (96%). This is summarised in Table 36: Number of patients assessed at 1 and 2 years. Our model showed that there were no statistically meaningful results or conclusions that can be drawn. This is most likely due to the limited sample size compared to the number of variables used. This will be re-run when more patients in the PITMS study reach their study visit timelines and more cohorts will be analysed in due course.

Table 45: Ocrelizumab predictors for NEDA at 12 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	1.82	1.53	1.20	0.23
Age at DMT initiation	0.10	0.04	2.21	**0.03***
Sex: Male	0.40	0.63	0.63	0.53
Ethnicity: Black	-1.61	1.13	-1.43	0.15
Ethnicity: Mixed	16.82	1569	0.01	0.99
Ethnicity: Other	-0.30	1.91	-0.15	0.88
Ethnicity: White	0.60	0.92	0.64	0.52
Phenotype: Optic neuritis	0.52	0.85	0.60	0.54
Phenotype: Spinal cord	-1.01	0.73	-1.37	0.17
Phenotype: Supratentorial	2.21	1.42	1.56	0.12
Presence of other comorbidities	-0.56	0.62	-0.90	0.37
No. of previous DMTs	-0.50	0.24	-2.03	**0.04***
Disease duration	-0.17	0.06	-2.91	**0.004****
No. of relapses in previous 12 months	0.70	0.40	1.73	0.08
No. of new MRI lesions at DMT initiation	-0.22	0.11	-1.90	**0.05***
EDSS at DMT initiation	-0.12	0.15	-0.82	0.41

Table 46: Ocrelizumab predictors for NEDA at 24 months				
	Coefficient (LogOdds)	**Std. Error**	**Z value**	**Pr(>|Z|)**
(intercept)	40.46	766	0.006	0.99
Age at DMT initiation	-0.01	0.05	-0.16	0.87
Sex: Male	1.22	0.97	1.25	0.21
Ethnicity: Black	-20.32	695	-0.00	0.99
Ethnicity: Mixed	-0.13	10000	0.00	1.00
Ethnicity: Other	-60.23	1299	-0.01	0.99
Ethnicity: White	-19.32	695	-0.00	0.99
Phenotype: Optic neuritis	-21.03	3209	-0.01	0.99
Phenotype: Spinal cord	-21.26	321	-0.01	0.99
Phenotype: Supratentorial	-21.54	3209	-0.01	0.99
Presence of other comorbidities	-0.70	0.83	-0.85	0.40
No. of previous DMTs	0.17	0.41	0.42	0.67
Disease duration	-0.01	0.01	-1.03	0.30
No. of relapses in previous 12 months	-0.32	0.50	-0.65	0.51
No. of new MRI lesions at DMT initiation	0.25	0.32	0.77	0.44
EDSS at DMT initiation	-0.17	0.21	-0.79	0.43

3.5 Discussion

In this chapter, we first described demographics, clinical features and NEDA outcomes in patients who have taken first- and second-line therapies at our centre. The basic demographics appeared to be similar in all cohorts including female: male ratio (approximately 3:1), ethnicity (white dominant), and age at CIS onset (late 20s and early 30s). Spinal cord clinical presentation at onset was the commonest phenotype. Disease duration was 5 – 7 years prior to starting therapy, with fingolimod reaching 9 years, likely because it is a second line therapy and usually given to patients who failed one or more first line therapies.

Most patients in our cohorts had low disability with median EDSS of 2.0-2.5, except in the natalizumab cohort where the median EDSS score was 6.0. This is likely because natalizumab is one of the highest efficacy therapies available for patients with highly active MS or those who have failed multiple other therapies with ongoing disease activity. Another observation is that many of our patients had no new MRI brain lesions at baseline, particularly so for first line therapies, but most patients had relapses in the previous 12 months of starting therapies. This can be explained for first line therapies as only one clinical relapse is required for commencing therapy. Another possible reason is that patients with spinal cord relapses may not have been accounted for through brain imaging. Finally, as expected, patients commenced on highly effective DMTs were more likely to achieve NEDA at 12 and 24 months compared to those initiated moderately effective DMTs.

The rate of achieving NEDA in our cohorts where comparable to other observational cohorts in the literature. The rate of NEDA in our DMF cohort was 76% and 58% at 1 and 2 years, respectively. Similarly, in a multicentre Italian cohort of 587 RRMS patients taking DMF, 80% achieved NEDA at 1 year and in a slightly larger retrospective American cohort of 737 RRMS patients, 61.2% reached NEDA at 2 years. (214, 215) In our fingolimod cohort, 61% achieved NEDA at 2 years, which was comparable (60%) to an Italian cohort of 201 RRMS patients. (216) The rate of NEDA in our natalizumab cohort was 69% at 2 years, which is similar to a smaller retrospective Norwegian cohort of 66 RRMS patients (68%). (217) Remarkably, 92%

of our ocrelizumab cohort achieved NEDA at 1 year, which is a similar finding (91.2%) reported in a Spanish retrospective multicentre cohort of 144 RRMS patients. (218) This rate dropped to 72.5% at 2 years, but remains higher than what was reported in an observational single centre cohort of 93 RRMS patients (62.1%). (219)

In the second part of this chapter, using frequentist analysis, we aimed to determine the impact of each variable, observed in our population, on the outcome, NEDA, by investigating each DMT cohort separately. We identified some statistically significant variables, mostly found in first line therapies, dimethyl fumarate and glatiramer acetate. Lower number of statistically significant variables were identified in the highly effective therapies, natalizumab and ocrelizumab, and hardly any in the fingolimod cohort. It is important to note however that both first-line therapies had the largest number of patients and the most variability (lower percentage) in achieving NEDA at 1 and 2 years, which usually helps predicting models.

Most of the statistically significant coefficients were negative, particularly in first line therapies, which in turn potentially indicates that such DMTs are more likely to work in patients with low MS activity. This reflects our current practice although highly effective therapies are also better used for those with low MS activity, they are associated with higher risks, such as PML with natalizumab, and are therefore usually reserved for patients with more MS activity. This is reflected with the positive coefficients of the number of relapses in the previous year of starting natalizumab and ocrelizumab and number of new MRI lesions at starting natalizumab and fingolimod.

It is also interesting to note that ethnicity appear to show no significance in achieving NEDA in any of the cohorts. It is widely observed in clinical practice and documented in literature that patients of African descent have greater baseline lesion burden, are at greater risk of disability progression and may not respond as well to some DMTs compared to patients of European descent. (220, 221) This is possibly due to our population being white dominant, reaching 78% in the GA cohort, which potentially reduces the impact of this variable.

Reviewing the baseline variables across all cohorts; older age at MS onset, lower number of previous DMTs, lower number of relapses in the previous 12 months, lower number of new and/or enhancing MRI brain lesions, and lower EDSS score appear to show significant association of achieving NEDA at 1 and 2 years for most cohorts. Lower number of previous DMTs was associated with NEDA in a cohort of RRMS patient taking ocrelizumab reported by Cellerino et al. (219) The number of relapses in the previous 12 months prior to starting fingolimod was found to be significantly associated with NEDA at 2 years by Giuliani et al. (216) Lower EDSS was also associated with reaching NEDA at 2 years in a cohort of 171 patients taking natalizumab, however age, previous relapses, and disease duration did not show significant associations. (222) Similarly, in the evaluation of a cohort of 215 RRMS patients, age, ethnicity, and disease duration at baseline, regardless of the type of DMT used, did not reach a statistical significance in reaching NEDA at 2 years. (223) Finally, in a large cohort (1032) of RRMS patient taking injectables (glatiramer acetate or interferons) 52.4% achieved NEDA at 2 years, and age and EDSS were significantly associated with reaching NEDA. (224)

Our study has identified a variety of baseline predictors of NEDA at 1 and 2 years which in most cases have been shown in separate other studies in the past years. It is also worth noting that sex and clinical phenotype at MS presentation appear to show no significant association. This could be due to the majority of MS patients being female and only a limited number of studies have considered the clinical phenotype at MS onset.

One of the main limitations to our study is that most of our data is retrospective and observational which was collected prospectively in a single centre. Ocrelizumab is the only prospective cohort in our population, which comprises 13% of the total data. All retrospective data was collected by me and was standardised so any variations in scoring that could potentially introduce bias, for example EDSS and relapses, were minimised. Larger data from existing international registries with greater number of patients and follow-ups could have greater power and generalisability. However, one of the commonest problems with international registries is the selection bias and the

neurologists' and patients' preference of certain DMTs. We have minimised such a problem by including all patients initiating DMTs in a single centre, where all patients who are going on treatment are discussed at our MDT clinics, thereby making treatment approach more homogeneous.

In our study we did not include enlarging lesions as per the NEDA definition. However, these are well known to be rare and unreliable, and most automatic lesion segmentation programmes, which are used in clinical trials to estimate lesion activity, report exclusively new lesions, which represent a more robust measure than enlarging lesions. It is likely that the definition of NEDA will evolve with technological innovations and clinical practice (225), and in the future it may include patient-related outcomes (PROMS) and brain atrophy, and drop enlarging lesions.

Although we included the number of prior DMTs taken before the index therapy as one of the variables, which can represent disease activity and/or intolerance, we did not include prior DMT potency and duration in our analysis. This can be included in future analysis to investigate the possible impact of such variables. Considering our median duration of being on DMTs is 3.8 years, we could explore longer-term NEDA, at 3 and 4 years, in future analysis and prediction.

It is difficult to apply similar conclusions on external cohorts in view of the outcomes being estimated, and the observed predictions are valid only for the range of data used to estimate the model. In addition to the known limitations of a linear regression, such as the sensitivity to outliers and the assumption of normality, there is no estimation of the uncertainty of the prediction. This is one of the limitations of standard, also known as frequentist, logistic regression models. This will be addressed in the next chapter which describes the use of Bayesian statistics to explore a different approach to explain the explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis. This is an essential step towards developing and validating a personalised tool which can potentially be used in routine clinical practice.

Chapter 4. Bayesian prediction of individualised treatment response in MS

4.1 Introduction

The disease course, mechanisms and treatment response of MS are heterogenous, which makes it challenging to predict outcome at the group-level, and even harder to predict outcome at the individual level. (199) But this task is important. A variety of DMTs have shown to improve clinical outcomes and quality of life at the group level, but if we are to pursue a personalised approach to DMT selection, we need the ability to forecast treatment response in any given *individual.*

Individual-level treatment responses can be forecasted by incorporating individual-level characteristics within predictive models of treatment effect. Accurate predictions will assist in choosing the correct therapy for an individual candidate and will minimise treatment failures and multiple therapy switches. Identifying the correct therapy for the right person at an early stage leads to reduced risk of relapses, accumulation of disability and adverse events. (226, 227)

Whilst factors including DMT type, number of relapses and new MRI lesions are associated with clinical response at the group-level, we do not have a tool that combines this information to make clinically useful predictions of treatment response at the individual-level. (228) Our objectives in this study are to predict NEDA, defined as no relapses, no MRI brain activity and no EDSS increase, in RRMS patients starting DMTs using routinely available clinical and radiological data, and a Bayesian modelling framework optimal for estimating individual-level uncertainty. This creates the basis for a future provision to support clinicians and patients when a treatment is needed for new RRMS patients and those who fail therapy and need to switch.

We applied a multi-level Bayesian framework which optimally estimates group-level and population-level coefficients in the context of data imbalance and limited participant subgroup numbers that typically occur in clinical settings.

4.2 Methods

4.2.1 Inclusion/Exclusion criteria

We analysed the final combination of our five cohorts described earlier (observational data from RRMS patients, aged ≥16 years old, who initiated glatiramer acetate, dimethyl fumarate, fingolimod, natalizumab or ocrelizumab between 2002–2022 at our centre). Patients started DMTs for PPMS or SPMS were excluded. We predicted NEDA (a binary variable) at 1 and 2 years from baseline, defined as the time of DMT initiation.

4.2.2 Predictor inclusion strategy and harmonisation

We included clinical and radiological predictors available to the MS neurologist in clinical practice including;

- DMT,
- sex,
- age at MS onset,
- ethnicity (grouped as White, Asian, Black, Mixed, or Other),
- the presence of one or more comorbidities,
- clinical phenotype at MS onset; characterised as optic neuritis (ON), supratentorial (ST), infratentorial (IT), spinal cord (SC) or other presentation (other),
- number of previous DMTs,
- The presence of alleles to HLA-DRB1*15
- The presence of unmatched OCBs in the CSF
- age at DMT initiation (baseline),
- disease duration from onset of CIS to commencing the reference DMT,
- number of relapses in the previous 12 months; a relapse is defined by new/worsening neurological symptoms lasting more than 24 hours in the absence of another cause such as an infection.
- EDSS at DMT initiation,

- brain MRI derived measures including; number of new T2w and/or T1w-gadolinium enhancing MS lesions and total lesion volume, which was added semi-automatically using ML algorithms.

All binary categorical variables were represented by single terms where 1 = presence and 0 = absence of the feature. DMTs, ethnicity and clinical phenotype at MS onset were all assigned into five categories each and so they were modelled as five individual categorical variables (represented as five terms for each variable). All retrospective data was collected by me and was standardised so any variations in scoring that could potentially introduce bias, for example EDSS and relapses, were minimised. Identifying patients through the pharmacy and MS speciality nurse databases reduced selection bias and using the electronic records have optimised our collection of data and minimised sampling bias.

4.2.3 MRI data acquisition and feature extraction

We acquired the MRI dataset from MS patients treated with DMTs, described above, according to local, standard-of-care protocols in the NHS over the last 20 years.

Baseline MRI features were collected from scans taken in the first 6 months of taking DMTs, unless scans from this time period were unavailable, in which case scans taken in the 6 months prior to initiating therapy were utilised. Where scans were not available or performed outside our centre, efforts were made to obtain scanned/copied documents to establish MRI findings described by neuro-radiologists.

The study had ethical and institutional approval at University College London Hospitals NHS Foundation Trust for the analysis of fully-anonymised, routinely collected data and was performed in accordance with the relevant guidelines and regulations.

We sought to extract the lesion load of each patient from their routine clinical MR imaging. This required registering all images to a common space and then segmenting

the lesions. In the retrospective cohorts, whole-brain FLAIR MR imaging of 1 x 1 x 6 mm resolution was acquired from a variety of sequences and parameters which had changed over time, and many different protocols were used. This may have introduced bias in terms of ability to identify lesions and measure volumes correctly. We used clinical neuro-radiologist reports for the number of lesions identified and used a semi-automatic lesion volume measure where poor quality scans were labelled manually. In the prospective PITMS ocrelizumab cohort, all MRI examinations were performed using a 3T MR system (Ingenia CX, Philips) and the product 16 channel neurovascular head coil, with an acquisition protocol including: a structural T1-weighted volume acquired using a 3D magnetisation prepared Turbo Field-Echo (3D-TFE) sequence with repetition time (TR) = 7 ms; echo-time (TE) = 3.2 ms; inversion delay time = 822 ms; flip angle = 8°; voxel size 1 x 1 x 1 mm^3, used as anatomical reference; 3D sagittal fluid-attenuated inversion recovery (3D-FLAIR) (TE=266 ms, TR=4800 ms, inversion delay time=1650 ms, flip angle=90°, 1.2x1.2x1.2 mm^3 resolution), used for lesion segmentation. The image pre-processing pipeline was assembled and executed by Dr Le Zhang (imaging-engineering post-doc research fellow). FLAIR images were denoised by the non-local means algorithm and rigidly registered to a T1-w image of isotropic 1mm resolution in MNI152 space. (229, 230) Images were skull-stripped using the Brain Extraction Tool (BET) and intensity corrected using the N4 algorithm. (231, 232)

A convolutional neural network (CNN) was trained to extract lesion volumes based on the framework for MS lesion segmentation published by Valverde et al in 2017. (233) Their training and inference procedures were followed, using 30 randomly selected manually labelled segmentations as the training set.

We conducted and reported this study according to STROBE (strengthening the reporting of observational studies in epidemiology) and a checklist has been provided in the appendix at the end of the thesis.

4.2.4 Model building

All analysis was done using RStudio v1.4.1106 (2021). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/). We used Bayesian multiple logistic regression models to develop a final 'comprehensive' model of individual prediction of NEDA. The Bayesian multiple logistic regression workflow was adapted from Jha et al. (234) NEDA (a binary variable) was entered as the dependent variable, whereas the other clinical and radiological variables were added as predictors. A Bayesian model is specified by the likelihood and priors.

1. $p(y_i = 1, X_i, Z_i, \alpha, \beta, \eta) \sim 1/(1 + \exp(-(\alpha + X_i\beta + Z_i\eta)))$ $for\ i = 1, \ldots, N$

2. $\beta_j \sim student\_t(df = 3, \mu_j = 0, \sigma_j = 2.5)$ $for\ j = 1, \ldots, J$

3. $\eta_{r,s} \sim normal(\mu_{r,s} = \pi, \sigma_{r,s} = \tau)$ $for\ s = 1, \ldots, S$; $for\ r = 1, \ldots, R$

4. $\pi_r \sim student\_t(df_r = 3, \mu_r = 0, \sigma_r = 2.5)$

5. $\tau_r \sim halfstudent\_t(df = 3, \mu = 0, \sigma = 2.5)$

6. $\alpha \sim N(\mu = 0, \sigma = 1)$

The logistic regression likelihood is shown in the first equation above, where $y_i = 1$ is the NEDA status for subject i of *N*. This is predicted by a transformed linear sum of three components $(\alpha + X_i\beta + Z_i\eta)$. In the first term α, is the estimated model intercept. The second term consists of a J-dimensional vector of observed *population-level* individual predictors, X_i, multiplied by a J-dimensional vector of estimated population-level regression coefficients β. Population-level variables included in X include continuous variables and binary variables and factors with 2 levels represented by single dummy variables where presence of the attribute was coded as 1. The third term, $Z_i\eta$, consists of an R x S dimensional vector of observed *group-level* individual predictors, Z_i, multiplied by an R (number of groups) x S (number of levels in each group) - dimensional vector of estimated group-level regression coefficients η. Group-level variables included in *Z* include factors or ordinal variables with 3 or more levels such as background ethnicity, EDSS and phenotype. The coefficient $\eta_{r,s}$ for each level *s* within group *r*, was further modelled with a

hierarchical distribution on the mean (π_r) and variance parameter (τ_r). This allows partial pooling of information from each level to optimise the variance of their coefficients and predictive performance of the posterior predictions. We followed a consistent and conservative approach to prior selection, employing weakly informative prior distributions over model parameters. (235) Uniform 'improper' priors were avoided as they are implausible and are liable to cause degeneracies in non-linear models. Student t distributions were preferred over Normal distributions where possible owing to their improved behaviour with the heavy tailed data characteristic of behaviour. (236) Priors over continuous unbounded population-level parameters of interest were generally modelled with a Student t distribution (degrees of freedom (df) = 3, mean = 0, sd = 2.5). Group-level intercepts were modelled hierarchically with normal priors whose mean and standard deviation were learned from the data (partial pooling) by placing Student t (df = 3, mean = 0, sd = 2.5) and Half-Student t (df = 3, mean = 0, sd = 2.5) priors over the Normal mean and standard deviations respectively.

Model Specification:

In order to determine the effect of increasing amounts of clinical information we specified and evaluated five comparator models of increasing complexity. The baseline (Null) model which had an intercept and no other predictors were used to estimate the average prediction of NEDA in our sample population. Further models were developed, as per Table 47: The Bayesian Models, cumulatively including additional features step-by-step: DMTs, History, Examination and Imaging.

Table 47: The Bayesian Models

Model Number	Model Name	Model Variables	No. of predictors
1	**Null**	Average population prediction	0
2	**+ DMTs**	Model 1 + DMTs	1
3	**+ History**	Model 2 + sex, age at DMT initiation, ethnicity, comorbidities, phenotype, disease duration, number of previous DMTs, number of relapses 12 months prior to DMT initiation	9
4	**+ Examination**	Model 3 + EDSS at DMT initiation	10
5	**+ MRI**	Model 4 + baseline number of new/enhancing MRI lesions, and total lesions volume	12

4.2.5 Posterior Estimation:

The posterior parameter estimates were generated using a Markov Chain Monte Carlo (MCMC) method. Analyses were performed with BRMS (237) and RStan (Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2. http://mc-stan.org/) running in RStudio v1.3.1056 (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/). RStan implements Hamiltonian Monte Carlo with No-U-Turn sampling. (238) This is a faster and more efficient way to explore the parameters space of complex models than other sampling methods, particularly when there are large numbers of correlated parameters. (234) This improved robustness and efficiency in terms of estimated sample size (ESS) of MCMC sample. Subsequently for each imputed dataset, 4 chains, each with 2000 iterations, were run from different starting values

randomly taken from a uniform distribution between -2 and +2. 1000 warm up samples were discarded followed by obtaining posterior estimates from a further 4000 samples (post-warmup) across all chains, which were assessed for chain stability and convergence using standard Rhat metrics within the software. (234)

Missing Data:

Predictor variables with >50% missing data were removed from the analysis; the presence/absence of alleles to HLA-DRB1*15 and presence/absence of unmatched OCBs. Other missing data were described in the previous chapter; Table 17: Missing data in the dimethyl fumarate cohort, Table 21: Missing data in the fingolimod cohort, Table 25: Missing data in the natalizumab cohort, Table 29: Missing data in the glatiramer acetate cohort.

Multiple imputation was used to impute the remaining missing data assuming that it was missing at random conditional on the other covariates measured, using MICE as previously described.

4.2.6 Model Evaluation:

The overall internal performance of all models was assessed with accuracy (number of correct predicted classifications/number of classifications) and balanced accuracy. Average internal discriminative performance was assessed with an Area Under the receiver-operator Curve (AUC).

Beyond internal validity, external validation (generalisability) is important to assess the model prediction and assess for 'over-fitting' when applied to new individuals. This ideally requires assessments of an external dataset from a different geographical population or time period. This is difficult to achieve at this moment, although we are in the process of validating our results in an external unseen cohort. In the meantime, we sought to perform leave-one-out cross-validation (LOO-CV) as a measure of generalisability. This estimates how well the model would perform on *new unseen individuals* from the population sampled. Bayesian models are computationally expensive to cross-validate and so we used an approximate technique based on Pareto smoothed importance-sampling. (239) This approximation can be assumed to be reliable as long as the Pareto tail parameter falls below 0.7. The resulting external validation metrics are based on a log-likelihood measure of probabilistic accuracy called the leave-one-out-expected log posterior density (elpd-loo). The final model comparison of how well the models generalise to new individuals was performed with the paired difference between their respective elpd-loo. Here, the worse model has a negative value relative to the preferred model which has a reference value and standard error of 0.

4.3 Results

4.3.1 Demographics and clinical outcomes

A total of 1986 patients (71% female; mean age 39.2 years (± 10.2); disease duration 8.4 years (± 7.0); median EDSS 2.5 [range 0 - 8.0]; 43% DMT naïve) initiated different DMTs (glatiramer acetate = 547 (27%); dimethyl fumarate = 670 (34%); fingolimod = 336 (17%); natalizumab = 177 (9%); ocrelizumab = 256 (13%)). The demographics and clinical information for each cohort are summarised in Table 48: Summary analysis of all cohorts.

Comparing cohorts, it is clear that in our population, the majority of patients were female and white. The median age of MS onset (or CIS) ranged between 27-30 years and the median age at starting DMT ranged between 36-40 years. The majority of patients presented with a spinal cord clinical phenotype at MS onset. There was a difference between the clinical characteristics of those who started first line versus second line therapies. Natalizumab patients, for example, showed higher rate of relapses, MRI activity and disability at baseline compared to patients who started first line therapies. This is expected considering natalizumab is usually reserved for highly active RRMS due to its high efficacy as well as balancing its important risk of PML. On the other hand, ocrelizumab patients appear to have less active MS, which again reflects the less restrictive use considering its high efficacy, more convenient mode of administration and frequency, and association with lower risks compared to natalizumab.

Overall, 1503/1986 (75.7%) of patients who commenced therapy achieved NEDA at year one. Only 1037/1835 (56.5%) patients achieved NEDA at year two. Patients included in NEDA analysis at each time point is summarised in Table 49: Number of patients assessed for NEDA at 1 and 2 years. As previously demonstrated in chapter three analysis, a small number of patients, 1-2% at year 1 and 2-4% at year 2, were lost to follow up in each cohort.

The results above demonstrate that despite the clinical guidance available for clinicians to commence different DMTs it is still very difficult to predict treatment response. This again highlights the importance of understanding baseline factors that could predict short- and medium-term outcomes to the individual patient.

Table 48: Summary analysis of all cohorts					
Cohort	**Glatiramer Acetate (547)**	**Dimethyl Fumarate (670)**	**Fingolimod (336)**	**Natalizumab (177)**	**Ocrelizumab (256)**
Gender F:M (F%)	407:140 (74)	483:187 (72)	252:84 (75)	114:63 (64)	160:96 (63)
Ethnicity					
White (%)	425 (78)	478 (71)	243 (75)	118 (67)	193 (75)
Black	33	29	16	20	16
Asian	43	65	35	16	27
Mixed	17	24	15	6	14
Other	15	22	12	11	6
Comorbidities					
1 = yes (%)	347 (63)	324 (48)	200 (60)	128 (73)	144 (56)
0 = no	186	346	134	48	112
Age at MS onset					
Median	30	30	28	27	30
Mean	31.4	32	28.6	27.7	31.7
Min	8	10	8	10	11
Max	67	73	56	65	65
Phenotype at MS onset					
SC = spinal cord (%)	210 (38)	236 (35)	106 (32)	41 (23)	102 (40)
ON = optic neuritis	128	155	73	38	61
ST = supratentorial	82	111	48	19	23
IT = infratentorial	98	124	76	37	60
Other		9	2	7	
OCBs (%)	191 (35)	221 (33)	120 (36)	61 (35)	
1 = Yes	175	206	112	58	No data
0 = No	16	15	8	3	
Genetics (%)	85 (16)	106 (16)	73 (22)	106 (60)	
0 = no alleles	7	9	19	28	
1 = one allele	15	13	8	1	No data
2 = two alleles	0	0	4	7	
Pending phenotype	63	84	42	57	
No. of previous DMTs					
0	342	335	12	26	138
1	150	250	186	91	54
2	48	62	104	38	40
3	6	15	32	19	15
4 or more	1	8	1	3	9
Age at DMT initiation					
Median	37	40	39	36	38
Mean	38.7	40.5	39.2	36	38.8
Min	16	16	15	15	18
Max	72	82	69	68	68
Disease duration					
Median	5	7	9	7	5
Mean	7.3	8.6	10.6	8.4	7.1
Min	0	0	0	0	0

Max	36	40	48	41	30
No. relapses in last 12 months					
0	149	236	86	13	75
1	251	308	160	52	120
2	116	104	74	73	53
3 or more	28	21	13	38	8
EDSS at DMT initiation					
Median	2.0	2.0	2.5	6.0	2.0
Mean	2.5	2.6	3.5	4.7	2.7
Min	0	0	0	1.0	0
Max	7.0	7.5	7.5	8.0	8.0
No. new MRI brain lesions					
0	187	438	231	60	119
1	41	73	31	20	39
2	40	55	15	25	27
3	31	49	10	24	13
4	28	20	8	15	16
5-9	36	19	12	16	19
10 or more	0	1	2	2	3
Total lesions volume (mL)	**Mean+/-SD**	**Mean+/-SD**	**Mean+/-SD**	**Mean+/-SD**	**Mean+/-SD**
Year 0	9304.8±11936.9	8927.6±10703.1	11087.8±12152.6	16146.7±14532.0	11933.9±13270.2
Year 1	7362.8±7677.8	9045.9±11247.1	10997.2±12368.7	14726.3±14092.4	11424.6±13724.2
Year 2	10781.2±12608.4	9599.3±11982.7	11610.9±12964.4	16357.2±15408.6	10697.0±13505.3
NEDA at 1 year					
1 = yes (%)	358/547 (65)	510/670 (76)	262/336 (78)	144/177 (81)	235/256 (92)
0 = no (%)	189/547 (35)	160/670 (24)	74/336 (22)	33/177 (19)	21/256 (8)
NEDA at 2 years					
1 = yes (%)	248/532 (46.6)	387/670 (58)	206/336 (61)	123/177 (69)	87/120 (72.5)
0 = no	284/532 (53.4)	283/670 (42)	130/336 (39)	54/177 (31)	33/120 (27.5)

Table 49: Number of patients assessed for NEDA at 1 and 2 years

	Glatiramer Acetate	Dimethyl fumarate	Fingolimod	Natalizumab	Ocrelizumab	Total number
Patients assessed for NEDA at 1 year	547	670	336	177	256	1986
Patients assessed for NEDA at 2 years	532	670	336	177	120	1835

4.3.2 Predictive performance

Internal validation:

All MCMC chains converged adequately with Rhat=1 for all relevant model parameters. Increasing model complexity by adding additional covariates improved internal (within-sample) predictive measures. The sensitivity and specificity of the models are shown across all thresholds as Area Under the Curve (AUC), as shown in Table 50: Bayesian models' internal evaluation

Table 50: Bayesian models' internal evaluation

Model		Model 1 (Null)	Model 2 (+DMTs)	Model 3 (+History)	Model 4 (+EDSS)	Model 5 (+MRI)
Year 1	**Accuracy**	0.81	0.81	0.81	0.82	0.82
	Balanced Accuracy	0.50	0.50	0.51	0.53	0.54
	AUC	0.50	0.62	0.68	0.72	**0.73**
Year 2	**Accuracy**	0.66	0.66	0.67	0.68	0.69
	Balanced Accuracy	0.50	0.50	0.54	0.56	0.57
	AUC	0.51	0.60	0.65	0.68	**0.69**

Cross-validation:

In order to determine whether the improved predictive performance was generalisable to unseen patients, we subsequently performed approximate cross-validated model comparison. The approximate leave-one-subject-out cross-validation technique was reliable with all observations being associated with a Pareto $K < 0.7$. This showed that the final 'comprehensive' model (+MRI) generalised to new individuals better than the simpler models.

Cross-validated prediction accuracy of the final 'comprehensive' model (reference ELPD-LOO = 0, SEM = 0) was substantially better than a simpler model of average treatment effect (ELPD-LOO = -41.7, SEM ± 10.2), and similar to a simpler model with clinical but no radiological information (ELPD-LOO = -2.6, SEM ± 2.9). A similar pattern was observed with NEDA at 2 years - the average treatment effect model performed worse (ELPD-LOO = -23.6, SEM ± 8.3) whilst the clinical information only without

imaging model was similar (ELPD-LOO = -1.0, SEM ± 2.2) to the 'comprehensive' model. This is further demonstrated in Table 51: Cross-validated prediction accuracy of the models and Figure 6: Approximate-Out of sample accuracy: ELPD-LOO for models predicting NEDA at 1 (A) and 2 (B) years.

Table 51: Cross-validated prediction accuracy of the models

		Year 1		Year 2	
Model no.	**Model name**	**ELPD-LOO**	**SEM**	**ELPD-LOO**	**SEM**
5	**+MRI**	0	0	0	0
4	**+EDSS**	-2.6	±2.9	-1.0	±2.2
3	**+History**	-28.2	±7.9	-12.7	±5.6
2	**+DMTs**	-41.7	±10.2	-23.6	±8.3
1	**Null**	-72.2	±13.1	-46.6	±11.0

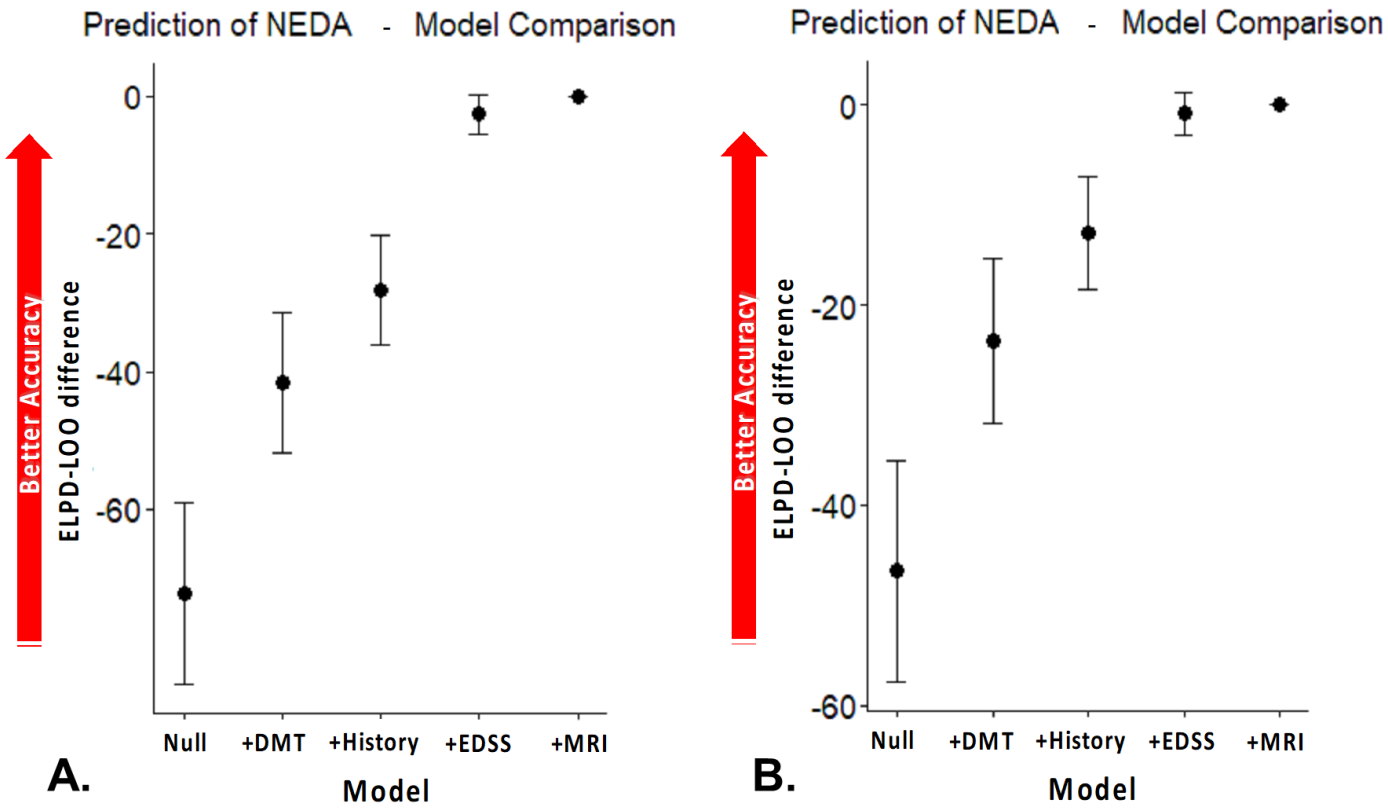


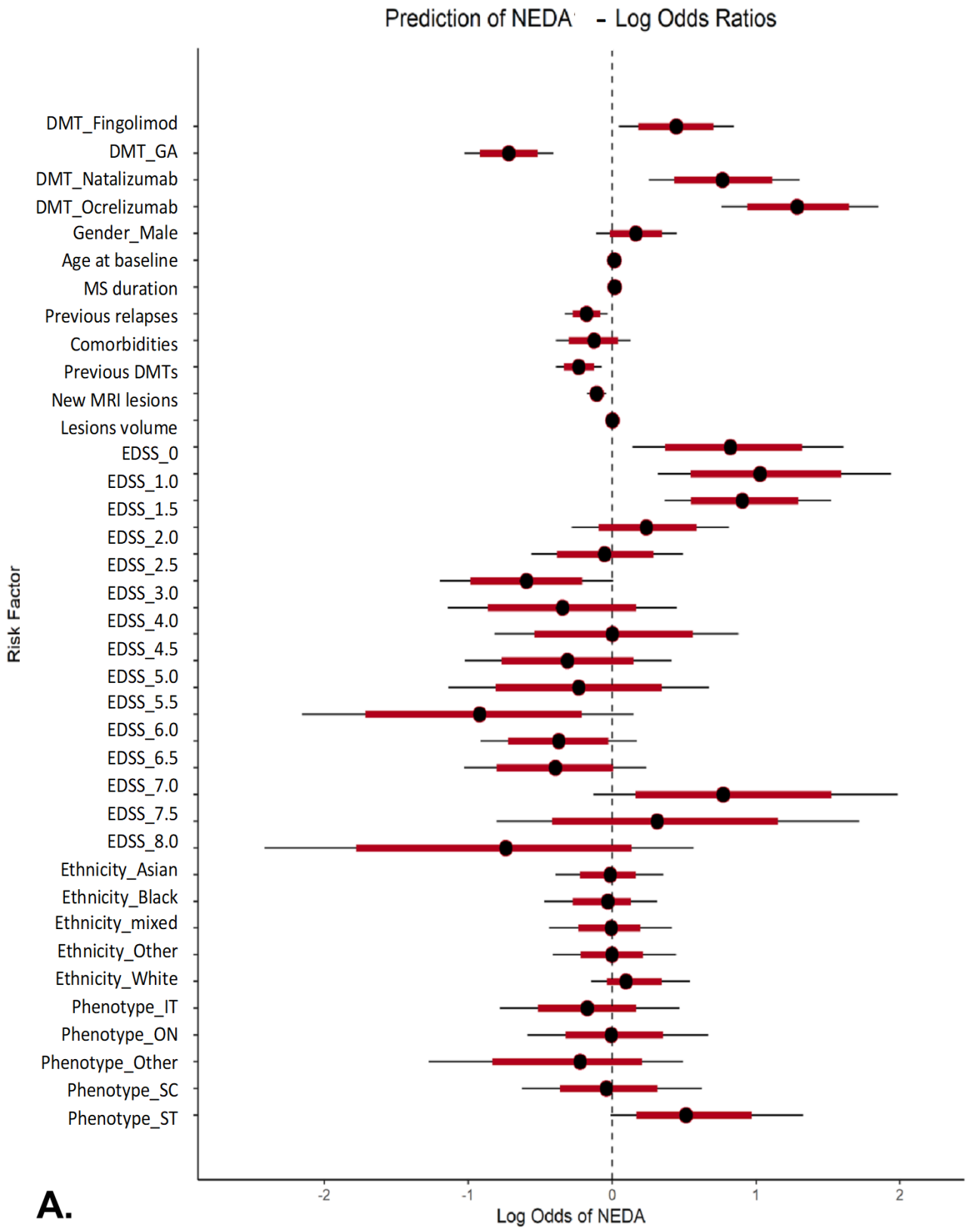
Figure 6: Approximate-Out of sample accuracy: ELPD-LOO for models predicting NEDA at 1 (A) and 2 (B) years.

4.3.3 Inference

The adjusted posterior log odds ratios of the fitted logistic regression model are presented in

Figure 7: The impact of variables in the most comprehensive model in predicting NEDA at 1 (A) and 2 (B) years

Bayesian models have no arbitrary significance threshold, so we assessed the strength of associations based on the mean effect-size (log odds ratio) and surrounding uncertainty. Values less than 0 are associated with a reduced probability of NEDA, whilst values greater than 0 are associated with an increased probability of NEDA, relative to the average population sample (black circle = average, red line = 80% Credibility Interval, black line 95% Credibility Interval). It is important to note that several variables were strongly associated with NEDA however, the observational nature of evaluating clinical cases limit causal interpretations.


Prediction of NEDA - Log Odds Ratios
DMT_Fingolimod
DMT_GA
DMT_Natalizumab
DMT_Ocrelizumab
Gender_Male
Age at baseline
MS duration
Previous relapses
Comorbidities
Previous DMTs
New MRI lesions
Lesions volume
EDSS_0
EDSS_1.0
EDSS_1.5
EDSS_2.0
EDSS_2.5
EDSS_3.0
EDSS_4.0
EDSS_4.5
EDSS_5.0
EDSS_5.5
EDSS_6.0
EDSS_6.5
EDSS_7.0
EDSS_7.5
EDSS_8.0
Ethnicity_Asian
Ethnicity_Black
Ethnicity_mixed
Ethnicity_Other
Ethnicity_White
Phenotype_IT
Phenotype_ON
Phenotype_Other
Phenotype_SC
Phenotype_ST
Risk Factor
-2
-1
0
1
2
Log Odds of NEDA

A.

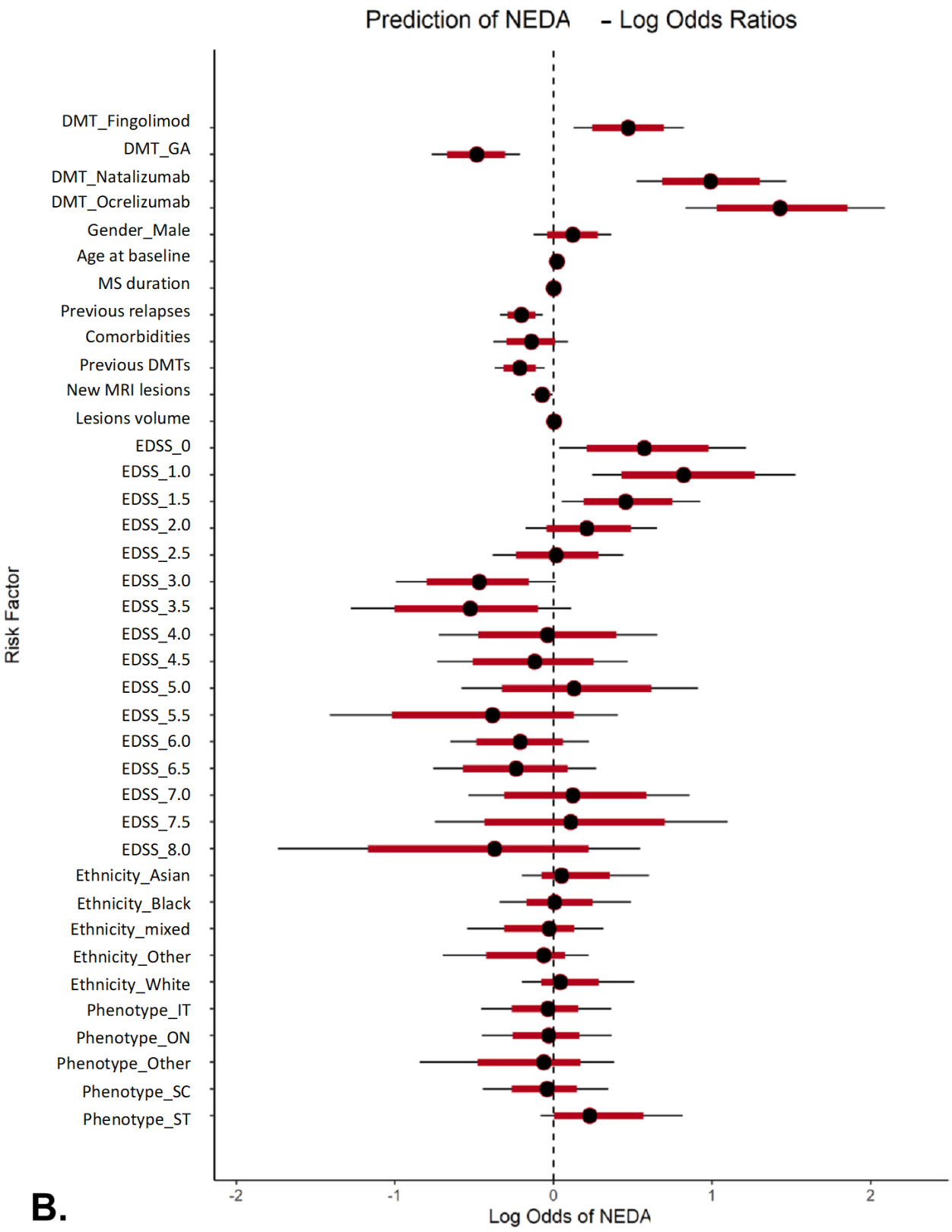


Figure 7: The impact of variables in the most comprehensive model in predicting NEDA at 1 (A) and 2 (B) years

4.4 Discussion

We developed and tested a modelling framework designed to answer a distinct question. We aimed to provide individualised predictions of treatment effects. We found that our model predicted individual NEDA at 1 and 2 years with reasonable accuracy based on the inclusion of all variables that are potentially available routinely to the MS neurologist. Although the results of cross-validation were consistent with this and encouraging, the use of the model in non-represented populations should still be cautious.

Previous work by Kalincik et al developed models to predict individual treatment response to seven DMTs using demographic, clinical and paraclinical predictors in 8513 patients with MS in the MSBase. (207) They assessed disability progression, disability regression, relapse frequency, and conversion to SPMS separately, using multivariable survival and generalised linear models. Although the accuracy and internal validity of the predictive models was high (>80%) for relapses at one year, it was moderate for relapses at 2-4 years and low for conversion to SPMS. However, this study did not consider radiological information and lacked data on dimethyl fumarate and ocrelizumab which are commonly used worldwide. In addition, assessing relapses and disability separately is important, however assessing NEDA is a more comprehensive measure of DMT effectiveness clinically and radiologically, and it is increasingly being used as primary outcome in clinical trials. (240) To date, to the best of our knowledge, there has been no attempt at predicting individualised NEDA in the short- or long-term using large real-world data.

We demonstrated that making our models more complex with the addition of further clinical and radiological variables helped in the improvement of prediction accuracy. We aimed to include other markers that have prognostic values, such as the presence of unmatched OCBs in the CSF and the presence of alleles to HLA-DRB1*15, but there was >50% of missing data and were therefore excluded in this analysis. (241, 242) However, we look to include more DMT cohorts, such as cladribine, alemtuzumab and beta-interferons, and more MRI variables, including white and grey matter volumes, in future analysis. This is one of the main advantages of using the Bayesian

predictive modelling framework as it is well setup to include newly available data and models can be monitored and updated regularly.

Our results have important implications for clinicians considering starting DMTs for RRMS patients in clinic, as there are two groups of DMTs available; moderately effective immunomodulating DMTs which are first line therapies (e.g interferon β, glatiramer acetate, dimethyl fumarate) and highly effective immunosuppressive and immune cell depleting monoclonal antibodies which are second line DMTs (e.g fingolimod, cladribine, natalizumab, ocrelizumab and ofatumumab). (243) According to the patient's MS activity, usually guided by the number of relapses, EDSS and MRI activity, the clinician decides upon the group of DMTs to commence. Importantly, our models have demonstrated that DMTs, as expected, have variable impact on prediction which correlate with their efficacy, however importantly the addition of further clinical and radiological information improved prediction further. This is an important finding for the treating clinician who wishes to commence a new DMT for a RRMS patient.

We chose to predict NEDA at one and two years as a measure of short- and intermediate-term treatment response respectively, similar to the follow up periods usually conducted by clinical trials. (110-112, 114, 115, 244, 245) Since we assessed both first- and second-line therapies together and first-line therapies take several months before they have a therapeutic effect, a period of two years to measure treatment effectiveness was thought to be appropriate. (246) On the other hand, second line-therapies are more potent and have a faster rate of effectiveness, but natalizumab, for example, is usually given for two years before it is switched to another DMT, particularly for those who are JC virus positive, to reduce the risk of developing PML. (247)

The Bayesian approach in prediction was previously used in epilepsy, where Jha et al used Bayesian models to develop and validate an individualised prediction of SUDEP (sudden unexpected death in epilepsy) risk in epilepsy patients using clinical predictors, with accuracy reaching 71%. (234) Their model also identified prognostic

factors at epilepsy onset including generalised tonic-clonic and focal-onset seizure frequency, young age and family history of epilepsy, which are important findings to explore by the treating clinician.

One of the main limitations to our study is the retrospective nature of our data. However, all data was collected in a standardised manner by a single operator (myself), so any variations in scoring, for example EDSS score and relapses, were minimised. Another limitation includes the lack of external independent cohort to test our model, however we are currently addressing this through ongoing collaboration with the Swedish MS team to test our model in an independent external cohort.

The Bayesian approach may be criticised because of the need for prior distributions for the model parameters. We therefore used weakly informative priors consistent with expert knowledge that standardised odds ratios in clinical studies are unlikely to be more or less than 2. (234)

Although the prediction remains – and will always remain - uncertain, we have developed a framework that potentially has some utility in clinical and research settings. Future prospectively acquired longitudinal PITMS study will provide more detailed cohorts with further information including various other variables that could potentially improve our model including genetics, serum neurofilament biomarkers, and further MRI characteristics.

Chapter 5. Comparing clinical and radiological effectiveness of disease modifying therapies in RRMS

5.1 Introduction

In the previous chapters we explored our experience in using different disease modifying therapies (DMTs) to manage RRMS. We evaluated treatment response according to NEDA (no evidence of disease activity) at one and two years through evaluating clinical (relapses and EDSS progression) and radiological activities. We also used machine learning to evaluate important clinical predictors of outcomes at a group level and at an individual level. In this chapter, we explore the clinical and radiological effectiveness of different DMTs given for RRMS patients.

In separately conducted clinical trials, all currently approved DMTs for RRMS have shown evidence of reducing relapses and progression at variable rates, as demonstrated in chapter one, Table 5: DMTs; effectiveness, mode of administration, monitoring, and adverse events. (111, 112, 114, 115, 245, 248) However, there has been no head-to-head phase III clinical trials comparing the efficacy of multiple first line DMTs and/or multiple high efficacy DMTs. In indirect comparisons with matching-adjusted analyses of clinical trials comparing two first line DMTs, peginterferon beta-1a showed better clinical outcomes when compared to subcutaneous interferon beta-1a, glatiramer acetate and teriflunomide. (249-251)

Real-world data (RWD) has become a valuable tool for the evaluation of clinical practices in the management of chronic conditions such as MS. (252, 253) Using RWD in the comparative effectiveness research analyses can guide healthcare systems, approval of therapies, and most importantly patient care and management. Recently Reder et al. used real-world propensity score comparison of treatment effectiveness of multiple first line (moderately effective) DMTs. (254) However, to date there has been a limited number of studies comparing multiple DMTs across lower, moderate and higher efficacy therapies in the literature. (207) We therefore conducted a propensity score-adjusted comparison of the clinical and radiological effectiveness of a combination of five low-, moderate- and high-efficacy DMTs among RRMS patients

at our tertiary MS centre. We conducted and reported this study according to STROBE (strengthening the reporting of observational studies in epidemiology) and a checklist has been provided in the appendix at the end of the thesis

5.2 Methods

5.2.1 Inclusion/Exclusion criteria

This observational study included all consecutive participants aged ≥16 years old who had a diagnosis of RRMS and initiated ocrelizumab, natalizumab, fingolimod, dimethyl fumarate (DMF) and glatiramer acetate (GA) at Queen Square MS Centre at the National Hospital for Neurology and Neurosurgery between 2002 and 2022. Patients with PPMS or SPMS on DMTs were excluded.

5.2.2 Data collection

A review of the electronic medical records was conducted for all the patients who were included in the study. I reviewed all the medical encounters following each patient's treatment initiation for at least 24 months or until DMT discontinuation. Identifying patients through the pharmacy and MS speciality nurse databases reduced selection bias and using the electronic records have optimised our collection of data and minimised sampling bias. Baseline information was collected from records at the time of starting a DMT. Baseline MRI characteristics were collected from the nearest scan after starting a DMT, this was usually within 6 months of initiating therapy. In the minority of patients who did not have available scans within this timeframe, scans undertaken in the 6 months prior to starting therapy were utilised.

5.2.3 Outcome measures

The primary outcome was comparing the clinical effectiveness of all DMTs defined as the patient experiencing a clinical relapse, EDSS progression and/or developing new T2/gadolinium (Gd) enhancing MRI brain lesions. We also compared NEDA (no evidence of disease activity), defined as the absence of relapses, EDSS progression and new MRI brain lesions (T2 and/or Gd enhancing MS lesions) at one and two years.

Since NEDA evaluates the three components in this analysis, the term NEDA will imply NEDA-3 in this chapter. For NEDA assessments, only those who had therapy for ≥6 months and ≥18 months were assessed for NEDA at 1 and 2 years, respectively.

5.2.4 Statistical analysis

Demographic, clinical, and MRI variables are presented as mean (±standard deviation), median (range), or number (percent), as appropriate. Using baseline variables, propensity scores were calculated for the probability of being assigned to DMTs based on age, gender, previous DMT, comorbidities, disease duration, EDSS, and relapses in previous 12 months.

Multivariable logistic and linear regression models were used to evaluate differences in probability and number of relapses, respectively, between different DMTs. Mixed-effect linear regression models were used to evaluate the differences in EDSS between baseline and after year 1 and 2, separately, between different DMTs, including an interaction term between time (year) and DMTs to explore between-DMTs differences over time. Multivariable logistic and linear regression models were used to evaluate differences in probability and number of new T2 and/or gadolinium-enhancing lesions, respectively, between different DMTs. Cox proportional hazards regression models were used to evaluate differences in NEDA between different DMTs. For the subgroup of patients with quantitative MRI analysis, mixed-effect linear regression models were used to evaluate the differences in lesion volume between baseline and after year 1 and 2, separately, between different DMTs, including an interaction term between time (year) and DMTs to explore between-DMTs differences over time. Ocrelizumab was used as reference DMT; propensity score was included as covariate in all statistical models. The proportional hazards assumption was met as assessed using plots of log (−log survival time) against log survival time and Schoenfeld residuals against survival time; linear regression of Schoenfeld residuals on time was used to test for independence between residuals and time.

Results were presented as odds ratio (OR), coefficient (Coeff), hazard ratio (HR), 95% confidence intervals (95%CI), and p-value, as appropriate. All P values were from 2-sided tests and results were deemed statistically significant at $P < 0.05$. Stata, version 15 MP (StataCorp LLC) was used to conduct statistical analyses. This statistical analysis was assisted by Dr Marcello Moccia, Assistant Professor at the University of Naples Federico II and Clinical Research Fellow at University College London, Dr Raffaele Palladino, clinical epidemiologist and statistician at school of public health, Imperial college London and Professor Maria Pia Sormani, Professor of biostatistics at the University of Genoa, Italy.

5.3 Results

5.3.1 Study population

We included 1986 RRMS patients (age=39.1±10.2 years; females=71.2%; disease duration=8.3±7.0 years; EDSS=2.5 (0-8.0)), who initiated ocrelizumab (256), glatiramer acetate (547), dimethyl fumarate (670), fingolimod (336) or natalizumab (177), with average duration on DMTs of 3.8±3.3 years.

For all patients, brain MRIs were acquired as per clinical practice, and we extracted presence and number of new T2 and Gd enhancing lesions. For a subset of patients (1600, 80.5%), we also had MRI acquisitions suitable for quantitative analysis, including white matter lesion volume at baseline and at least at one follow-up.

Demographic, clinical, and MRI variables are presented in Table 52: Demographic, clinical, and MRI variables at baseline, year 1, and year 2 for different DMTs. Many of the variables mentioned in this table were also described in the previous chapter, Table 48: Summary analysis of all cohorts. Important difference is that in the analysis in comparing DMTs, we excluded patients who received therapy for <6 months and <18 months for the assessment at 1 and 2 years, respectively. A summary of the number of patients assessed at 1 and 2 years is shown in chapter three, Table 36: Number of patients assessed at 1 and 2 years. Reasons for patients stopping DMTs at different time points, within 2 years, were discussed in chapter 3.

Table 52: Demographic, clinical, and MRI variables at baseline, year 1, and year 2 for different DMTs.					
	Ocrelizumab N=256	**Dimethyl fumarate N=670**	**Fingolimod N=336**	**Glatiramer acetate N=547**	**Natalizumab N=177**
Baseline					
Age, *years*	38.8±10.6	40.5±10.2	39.2±10.4	38.6±9.5	36.1±10.2
Gender, *females*	160 (62.5%)	483 (72.0%)	252 (75.0%)	407 (74.4%)	113 (63.8%)
Comorbidities	144 (56.2%)	324 (48.3%)	200 (60%)	347 (63.4%)	131 (74.0%)
Disease duration, *years*	7.0±6.8	8.6±7.1	10.6±7.1	7.2±6.6	8.4±6.6
EDSS	2.0 (0-8.0)	2.0 (0-7.5)	2.5 (0-7.5)	2.0 (0-7.0)	6.0 (1.0-8.0)
Average number of relapses in 12 months before treatment	0.97±0.79	0.86±0.78	1.05±0.83	1.04±0.84	1.85±1.03
Treatment naïve patients	138 (53.9%)	335 (50.0%)	12 (3.6%)	342 (62.5%)	26 (14.7%)
Availability of quantitative MRI	162 (63.2%)	621 (92.6%)	317 (94.3%)	338 (61.7%)	162 (91.5%)
Lesion volume*, mL*	11933.9±13270.2	8927.6±10703.1	11087.8±12152.6	9304.8±11936.9	16146.7±14532.0
Year 1					
EDSS	2.0 (0-8.5)	2.0 (0-7.5)	2.5 (0-7.5)	2.0 (0-7.5)	6.0 (1.0-8.5)
Patients with relapses	3 (1.1%)	75 (12.3%)	50 (15.7%)	128 (24.9%)	27 (15.2%)
Number of relapses	0.01±0.10	0.14±0.42	0.20±0.52	0.29±0.55	0.20±0.53
Patients with new/Gd+ lesions	14 (7.3%)	47 (8.8%)	12 (5.0%)	46 (25.1%)	11 (6.5%)
Patients with NEDA (baseline to year 1)	235/255 (92.1%)	510/610 (83.6%)	262/317 (82.6%)	358/514 (69.6%)	144/177 (81.3%)
Months to loss of NEDA (baseline to year 1)	11.6±1.4	11.4±1.9	11.5±1.7	11.1±2.4	11.7±1.3
Lesion volume*, mL*	11424.6±13724.2	9045.9±11247.1	10997.2±12368.7	7362.8±7677.8	14726.3±14092.4
Lesion volume*, percent change (baseline to year 1)*	3.66±49.6%	13.56±61.79%	11.99±82.66%	17.70±62.53%	0.86±43.57%
Year 2					
EDSS	2.5 (0-8.0)	2.0 (0-7.5)	2.5 (0-7.5)	2.0 (0-8.0)	6.0 (1.0-8.5)
Patients with relapses	6 (5.21%)	57 (10.9%)	48 (16.8%)	85 (21.0%)	18 (10.6%)
Number of relapses	0.05±0.23	0.15±0.65	0.18±0.43	0.22±0.45	0.12±0.38
Patients with new/Gd+ lesions	3 (4.1%)	61 (13.7%)	16 (7.1%)	37 (27.8%)	6 (3.5%)
Patients with NEDA (baseline to year 2)	87/115 (75.7%)	387/547 (70.7%)	206/299 (68.9%)	248/458 (54.1%)	123/169 (72.8%)
Months to loss of NEDA (baseline to year 2)	22.8±4.3	21.3±5.8	21.3±5.7	19.7±6.9	21.4±5.4
Patients with NEDA (year 1 to year 2)	91 (88.3%)	407 (79.1%)	222 (77.8%)	280 (70.0%)	143 (84.6%)
Months to loss of NEDA (year 1 to year 2)	23.8±0.7	23.8±1.1	23.8±1.0	23.6±1.4	23.9±0.7
Lesion volume*, mL*	10697.0±13505.3	9599.3±11982.7	11610.9±12964.4	10781.2±12608.4	16357.2±15408.6
Lesion volume*, percent change (baseline to year 2)*	-3.64±30.31%	27.26±102.77%	18.30±50.95%	16.43±66.00	3.90±49.70%

5.3.2 Relapses

On year 1, propensity score-adjusted logistic regression models showed a higher probability of relapse in patients treated with dimethyl fumarate (OR=11.73; 95%CI=3.65, 37.72; p<0.01), fingolimod (OR=15.98; 95%CI=4.91, 51.97; p<0.01), glatiramer acetate (OR=28.63; 95%CI=8.98, 91.25; p<0.01), and natalizumab (OR=15.23; 95%CI=4.54, 51.07; p<0.01), when compared with ocrelizumab (Figure 8: Bar graph showing the percent of patients with relapses during 12 months before treatment, and after 1 and 2 years.). Similarly, at year 1, propensity score-adjusted linear regression models showed a higher number of relapses in patients treated with dimethyl fumarate (Coeff=0.13; 95%CI=0.06, 0.20; p<0.01), fingolimod (Coeff=0.19; 95%CI=0.11, 0.26; p<0.01), glatiramer acetate (Coeff=0.28; 95%CI=0.21, 0.36; p<0.01), and natalizumab (Coeff=0.19; 95%CI=0.10, 0.28; p<0.01), when compared with ocrelizumab.

On year 2, propensity score-adjusted logistic regression models showed a higher probability of relapse in patients treated with fingolimod (OR=3.26; 95%CI=1.34, 7.90; p<0.01), and glatiramer acetate (OR=4.75; 95%CI=2.00, 11.26; p<0.01), when compared with ocrelizumab, while no differences were found for dimethyl fumarate (OR=2.12; 95%CI=0.88, 5.08; p=0.09), and natalizumab (OR=1.89; 95%CI=0.72, 4.96; p=0.19) (Figure 8: Bar graph showing the percent of patients with relapses during 12 months before treatment, and after 1 and 2 years.). Similarly, at year 2, propensity score-adjusted linear regression models showed a higher number of relapses in patients treated with fingolimod (Coeff=0.12; 95%CI=0.01, 0.24; p=0.03), and glatiramer acetate (Coeff=0.18; 95%CI=0.06, 0.29; p<0.01), when compared with ocrelizumab, while no differences were found for dimethyl fumarate (Coeff=0.10; 95%CI=-0.01, 0.21; p=0.07), and natalizumab (Coeff=0.06; 95%CI=-0.06, 0.19; p=0.30).

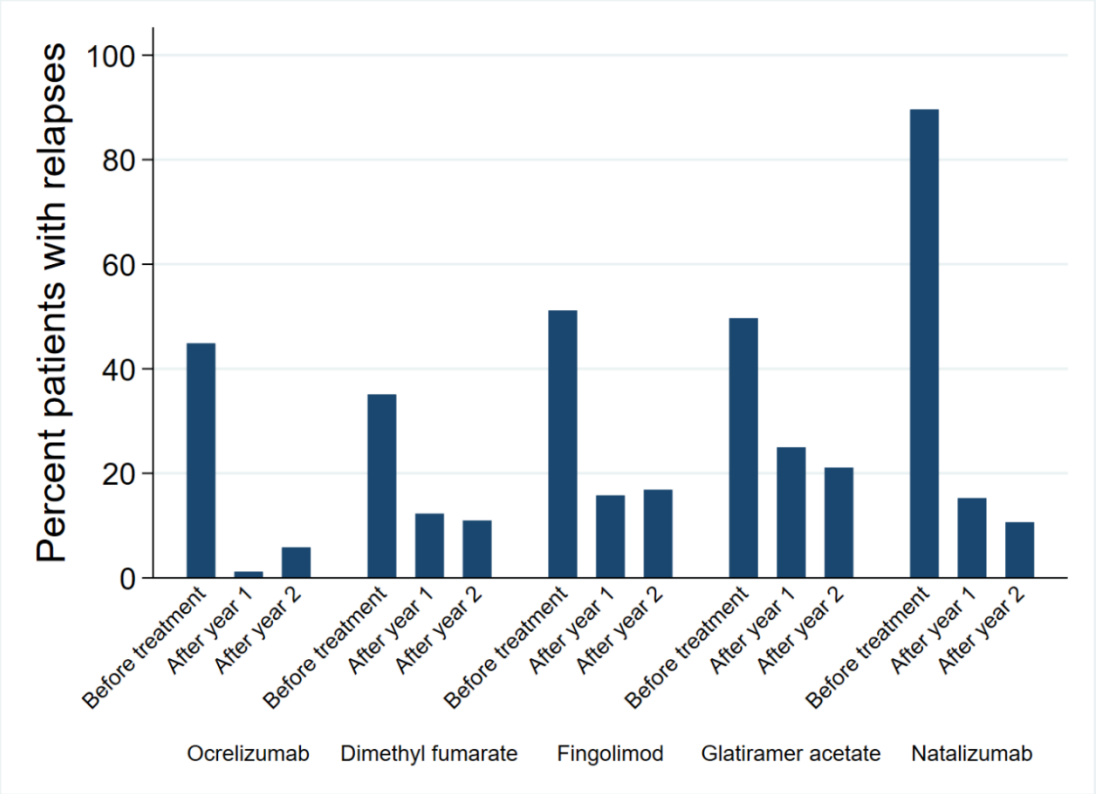


Figure 8: Bar graph showing the percent of patients with relapses during 12 months before treatment, and after 1 and 2 years.

5.3.3 EDSS

On year 1, propensity score-adjusted mixed-effect linear regression models also including baseline EDSS, showed EDSS was 0.24 larger in patients treated with dimethyl fumarate (Coeff=0.24; 95%CI=0.16, 0.32; p<0.01), 0.25 in fingolimod (Coeff=0.25; 95%CI=0.16, 0.35; p<0.01), 0.34 in glatiramer acetate (Coeff=0.34; 95%CI=0.25, 0.42; p<0.01), and 0.21 in natalizumab (Coeff=0.21; 95%CI=0.11, 0.32; p<0.01), when compared with ocrelizumab (Figure 9: Box-and-whisker plot showing EDSS before treatment, and after 1 and 2 years).

On year 2, propensity score-adjusted mixed-effect linear regression models, showed higher EDSS in patients treated with dimethyl fumarate (Coeff=0.22; 95%CI=0.08, 0.35; p<0.01), fingolimod (Coeff=0.24; 95%CI=0.10, 0.39; p<0.01), glatiramer acetate (Coeff=0.39; 95%CI=0.25, 0.53; p<0.01), and natalizumab (Coeff=0.21; 95%CI=0.06, 0.37; p<0.01), when compared with ocrelizumab (Figure 9: Box-and-whisker plot showing EDSS before treatment, and after 1 and 2 years).

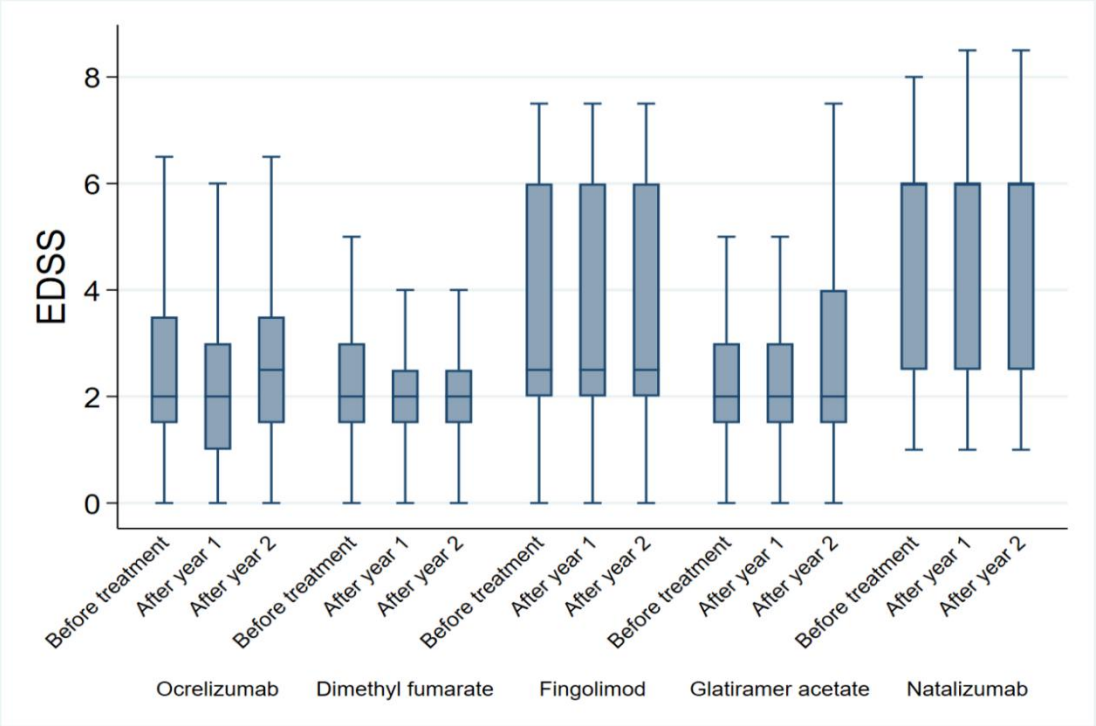


Figure 9: Box-and-whisker plot showing EDSS before treatment, and after 1 and 2 years.

The lines splitting the boxes in two represent the median. The top and bottom edges of the boxes represent the upper and lower quartiles respectively. The top and bottom of the straight lines represent the maximum and minimum values of the cohort. The figures are raw presentation of data.

5.3.4 MRI

On year 1, propensity score-adjusted logistic regression models showed a higher probability of new/Gd enhancing lesions in patients treated with glatiramer acetate (OR=4.07; 95%CI=2.03, 8.16; p<0.01), when compared with ocrelizumab, while no differences were found for dimethyl fumarate (OR=1.45; 95%CI=0.74, 2.84; p=0.27), fingolimod (OR=0.88; 95%CI=0.38, 2.02; p=0.77), and natalizumab (OR=0.84; 95%CI=0.36, 1.96; p=0.70) (Figure 10: Bar graph showing the percent of patients with new/Gd enhancing lesions at treatment initiation (baseline), and after 1 and 2 years.).

On year 2, propensity score-adjusted logistic regression models showed a higher probability of new/Gd enhancing lesions in patients treated with dimethyl fumarate (OR=4.93; 95%CI=1.45, 16.68; p=0.01), and glatiramer acetate (OR=9.40; 95%CI=2.67, 33.05; p<0.01), when compared with ocrelizumab, while no differences were found for fingolimod (OR=1.71; 95%CI=0.44, 6.53; p=0.43), and natalizumab (OR=0.77; 95%CI=0.18, 3.26; p=0.72) (Figure 10: Bar graph showing the percent of patients with new/Gd enhancing lesions at treatment initiation (baseline), and after 1 and 2 years.).

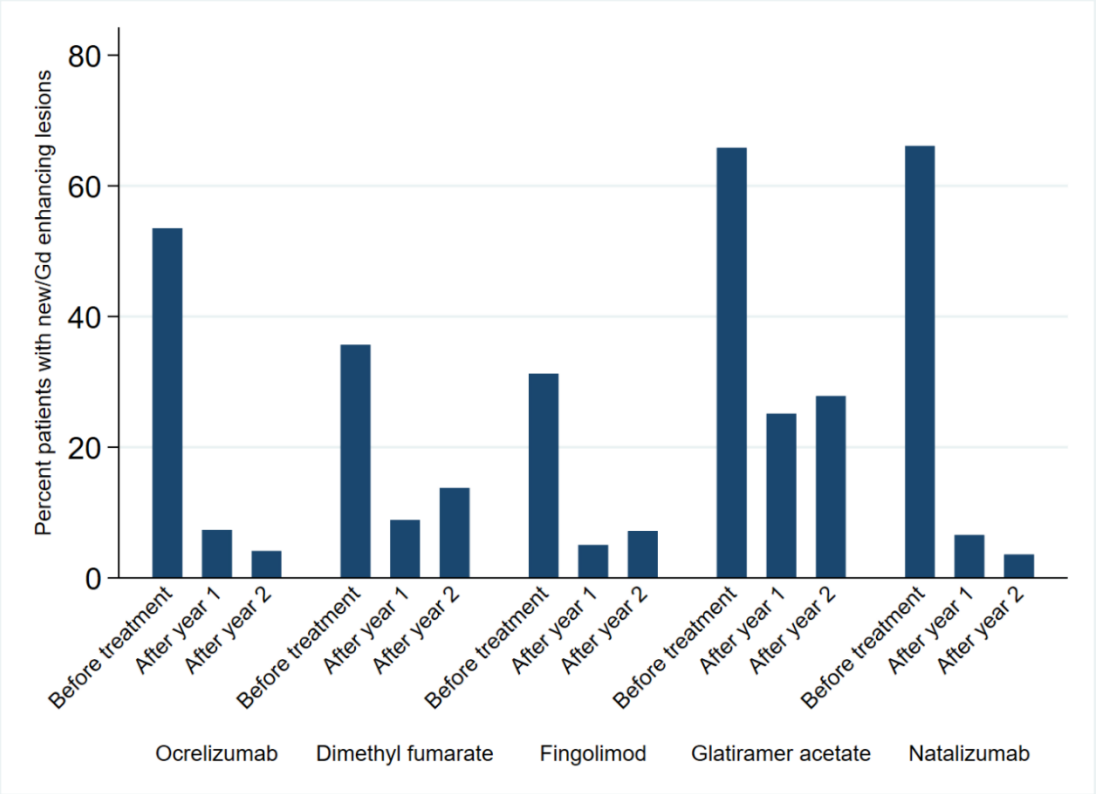


Figure 10: Bar graph showing the percent of patients with new/Gd enhancing lesions at treatment initiation (baseline), and after 1 and 2 years.

5.3.5 NEDA

On year 1, propensity score-adjusted Cox regression models showed reduced probability of remaining NEDA in patients treated with dimethyl fumarate (HR=0.49; 95%CI=0.36, 0.65; p<0.01), fingolimod (HR=0.54; 95%CI=0.38, 0.75; p<0.01), glatiramer acetate (HR=0.55; 95%CI=0.41, 0.73; p<0.01), and natalizumab (HR=0.40; 95%CI=0.27, 0.61; p<0.01), when compared with ocrelizumab (Figure 9a).

On year 2, propensity score-adjusted Cox regression models showed reduced probability of remaining NEDA in patients treated with dimethyl fumarate (HR=0.53; 95%CI=0.39, 0.72; p<0.01), fingolimod (HR=0.53; 95%CI=0.38, 0.74; p<0.01), glatiramer acetate (HR=0.66; 95%CI=0.49, 0.89; p<0.01), and natalizumab (HR=0.52; 95%CI=0.35, 0.77; p<0.01), when compared with ocrelizumab (Figure 9b).

Between year 1 and 2, propensity score-adjusted Cox regression models showed reduced probability of remaining NEDA in patients treated with dimethyl fumarate (HR=0.38; 95%CI=0.26, 0.55; p<0.01), fingolimod (HR=0.39; 95%CI=0.26, 0.59; p<0.01), glatiramer acetate (HR=0.48; 95%CI=0.33, 0.70; p<0.01), and natalizumab (HR=0.35; 95%CI=0.20, 0.60; p<0.01), when compared with ocrelizumab (Figure 9c).

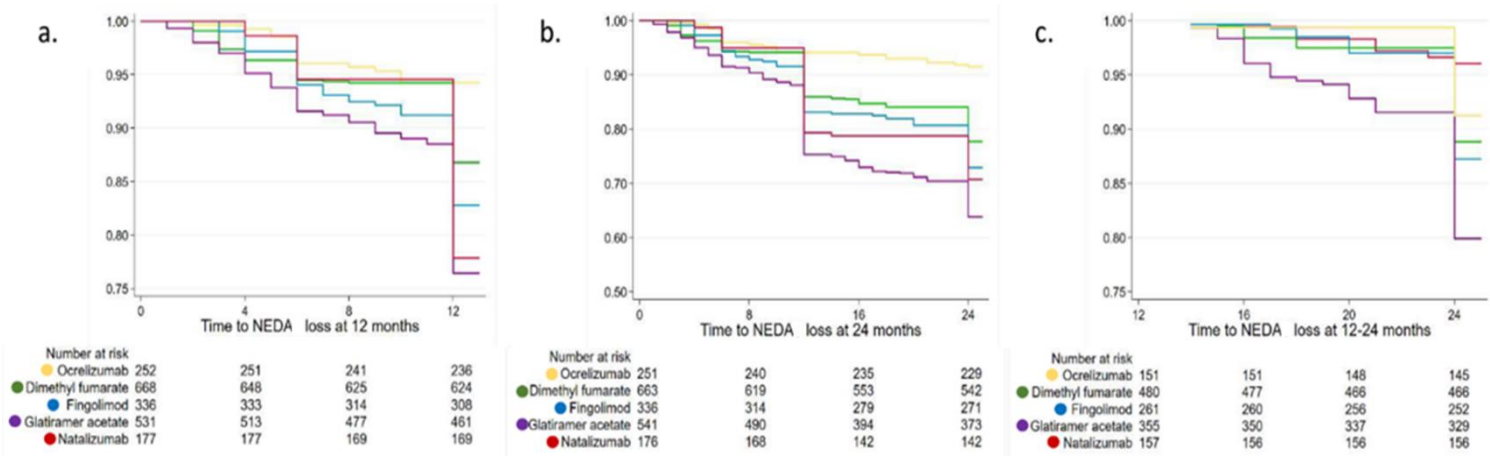


Figure 11: Kaplan-Meier curve showing the rate of NEDA loss for ocrelizumab (yellow), dimethyl fumarate (green), fingolimod (blue), glatiramer acetate (green), and natalizumab (red), after 1 year (a), 2 years (b), and between year 1 and 2 (c).

5.3.6 Lesion volume

On year 1, propensity score-adjusted mixed-effect linear regression models did not show any differences in lesion volume between patients treated with dimethyl fumarate (Coeff=219.43; 95%CI=-703.82, 1142.70; p=0.64), fingolimod (Coeff=127.79; 95%CI=-909.43, 1165.03; p=0.80), glatiramer acetate (Coeff=-85.94; 95%CI=-1371.88, 1199.99; p=0.89), natalizumab (Coeff=-927.77; 95%CI=-2059.44, 203.69; p=0.10), and ocrelizumab (Figure 12: Box-and-whisker plot showing lesion volume before treatment, and after 1 and 2 years).

On year 2, propensity score-adjusted mixed-effect linear regression models did not show any differences in lesion volume between patients treated with dimethyl fumarate (Coeff=2147.31; 95%CI=-145.96, 4440.58; p=0.06), fingolimod (Coeff=1773.48; 95%CI=-586.10, 4133.07; p=0.14), glatiramer acetate (Coeff=2112.52; 95%CI=-475.48, 4700.54; p=0.11), natalizumab (Coeff=636.10; 95%CI=-1789.19, 3061.40; p=0.60), and ocrelizumab (Figure 12: Box-and-whisker plot showing lesion volume before treatment, and after 1 and 2 years).

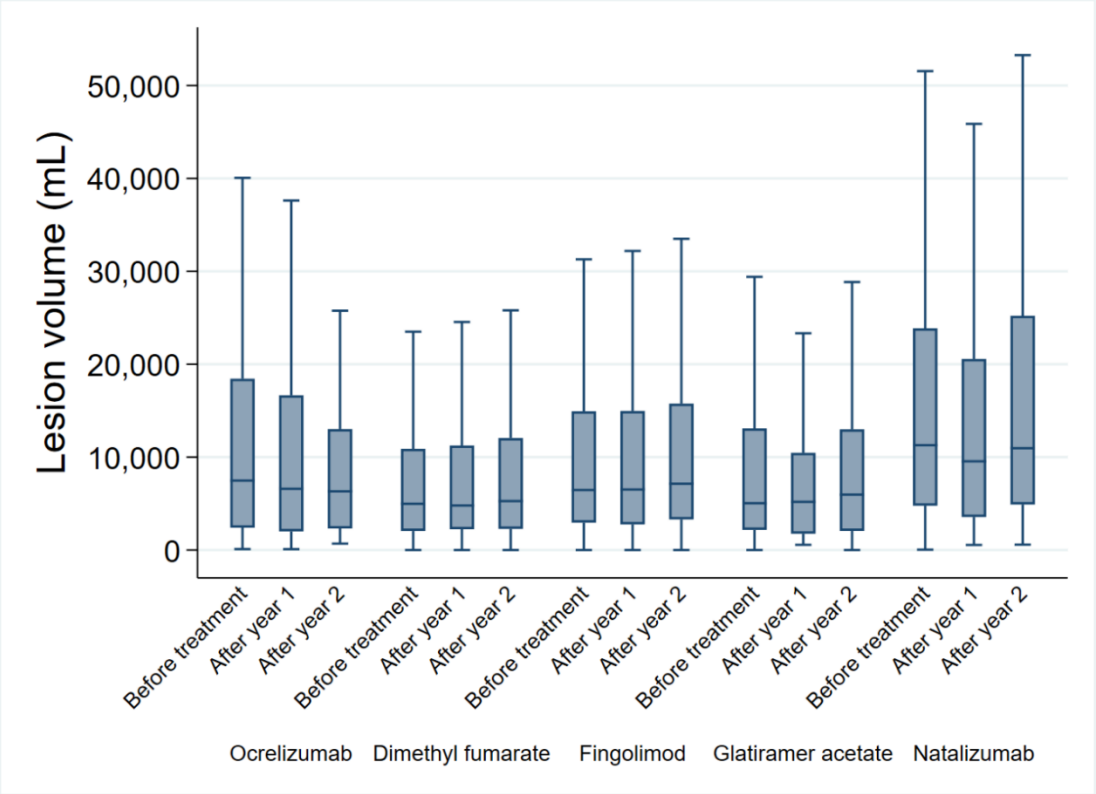


Figure 12: Box-and-whisker plot showing lesion volume before treatment, and after 1 and 2 years.

The lines splitting the boxes in two represent the median. The top and bottom edges of the boxes represent the upper and lower quartiles respectively. The top and bottom of the straight lines represent the maximum and minimum values of the cohort

5.4 Discussion

In this real-world observational study, we compared the effectiveness of five DMTs taken by RRMS patients in the past two decades. A large sample was obtained and a well-balanced cohort for comparison was achieved through propensity-score adjusting.

Our findings show that further real-world data (RWD) analyses of RRMS patients commencing different DMTs will likely compliment the evidence generated from randomized clinical trials. (255) This is because RWD can increase the understanding of the relative effectiveness of DMTs, long-term DMT efficacy and safety, treatment discontinuation and switching outcomes, and the impact of other clinical variables, such as age and comorbidities, on treatment effectiveness and disease progression. (255) We are looking to include more retrospective cohorts, including beta-interferons, alemtuzumab and cladribine, which will give us more information and power to our analysis. A previous small single-centre Canadian study showed limited difference in achieving NEDA between alemtuzumab and cladribine, but it would be interesting to see how they can be compared in our much bigger population. (256)

Our study demonstrated that patients taking ocrelizumab in our cohort achieved higher rate of NEDA status and had lower EDSS worsening compared to other DMTs, using propensity score adjustment. We also noted that ocrelizumab showed a slightly better outcomes compared to natalizumab, though this applies to propensity score adjusting and not matching. There has been no previous direct comparison between ocrelizumab and natalizumab, however, in a large retrospective analysis comparing the effectiveness of ocrelizumab, natalizumab and cladribine in RRMS patients switching from fingolimod, Zhu et al found that patients had lower annualised relapse rate for ocrelizumab compared to natalizumab. (257) However, statistical significance was not reached due to the low number of annualised relapse rate and limited period of observation in both groups.

Different propensity score methods can be used to attempt to mimic randomisation when comparing outcomes based on RWD sources. Some of the most common

methods are propensity score matching, adjusting and inverse probability of treatment weighting (IPTW). (258) We chose propensity score adjusting for our study considering our small number of patients and to generate a pilot data exploring this method of analysis. When using more robust analysis in performing propensity score matching, we used a multinomial logistic regression model to estimate the probability of treatment allocation of each DMT compared with ocrelizumab, based on age, gender, disease duration, comorbidities, previous DMTs, relapses in the previous year, EDSS and the year of DMT initiation. Based on propensity scores calculated, for each pair of DMTs (ocrelizumab vs dimethyl fumarate, ocrelizumab vs fingolimod, ocrelizumab vs glatiramer acetate, and ocrelizumab vs natalizumab), patients were matched on a 1:1 ratio, by nearest neighbour with replacement, within a calliper of 0.1 standard deviation of the propensity score. No patient satisfied matching criteria and no further analyses were run. Same results were obtained when we used four different logistic regression models to estimate the probability of treatment allocation of each DMT compared with ocrelizumab, separately (ocrelizumab vs dimethyl fumarate, ocrelizumab vs fingolimod, ocrelizumab vs glatiramer acetate, and ocrelizumab vs natalizumab), based on age, gender, disease duration, comorbidities, previous DMTs, relapses in the previous year, EDSS and year of treatment initiation. This shows the difficulty of matching patients in our cohort considering the constant development of therapies and evolving guidelines of treatment eligibility over time. Such common issue found in single centre studies can become less problematic when assessing much bigger cohorts in multicentre databases.

Although our study showed similar NEDA rates at 1 and 2 years compared to other observational studies, as discussed in chapter three, our NEDA rates are much higher compared to phase III trials. For example, our NEDA rate in patients taking ocrelizumab reached 75% at 2 years, whereas it only reached 48% at 2 years in a randomised control trial. (259) This higher rate is expected in real world analysis where clinicians are guided by regularity bodies, NHS England for example, and patients are stratified to different therapies according to results obtained from phase III trials to optimise care and provide the best treatment that suite the right patient's

disease characteristics. Besides, in prospective trials the follow-up periods, scanning and clinical reviews are much more frequent and detailed, and therefore more likely to detect disease activity compared to real world clinical setting.

The less striking results on MRI lesion volumes could be due to the limited sample size analysed. Further analysis on more MRI outcomes, such as white and grey matter volumes, is ongoing.

To the best of our knowledge, there has been no such sizable comparative study including such a variety of DMTs including ocrelizumab. Similar smaller studies compared different DMTs to rituximab, which is used more commonly in Europe as off-label therapy of MS. Vollmer et al used retrospectively collected data to compare the effectiveness of rituximab versus natalizumab, fingolimod and dimethyl fumarate using logistic regression on matched patients through propensity scores. (260) They found that patients taking rituximab had lower odds of experiencing disease activity compared to fingolimod and DMF, but similar odds to natalizumab.

Two other much smaller studies failed to show a difference between clinical relapses for natalizumab and rituximab after adjustment. (261, 262) Both studies only included clinical relapses and MRI outcomes, but a composite measure may be needed when comparing effectiveness between highly effective DMTs. This is why we also included NEDA in our study as it increases power to detect smaller differences through observation of more events. (64)

Since our study is observational, retrospective and non-randomised there were limitations. EDSS was our only measure of disability as this is usually performed and documented regularly in the MS clinic. It is widely used in clinical practice and research as it is easily performed and interpreted by clinicians. However, one of the main criticisms of EDSS is that it is regarded as heavily weighted on mobility and not very sensitive to upper limb disability and cognitive impairment. (52) Patient reported outcomes and the 9-hole-peg test, for example, have been used in MS research for the assessment of cognitive impairment and upper limb dysfunction, and would be

great to be included in future studies and assessments of treatment response. (54, 263)

Our cohort consists of mainly retrospectively collected data and a smaller proportion of prospectively collected data including the ocrelizumab cohort. This may have introduced bias as gathering data for patients who are in a study attending scheduled visits minimises missing information and assessments can be more accurately in a study setting rather than a clinic setting. This may contribute to the improved outcomes for ocrelizumab compared to other DMTs collected retrospectively. Besides, ocrelizumab is one of the latest approved highly effective DMTs and inclusion criteria has been less restrictive compared to other highly effective therapies such as natalizumab. This has made the cohorts evidently different in terms of level of MS activity, relapses and MRI lesions, and disability, including EDSS scores.

Different DMTs have been introduced to patients at different time periods over the last two decades and therefore some confounders are time varying and can be affected by prior treatment. Although we gathered data including number of previous DMTs used prior to the use of the index DMT, prior DMTs can affect confounders over time. Statistical methods, such as inverse probability weighting, can be used in future analysis to address this treatment confounder feedback. Also, we did not include the year of the baseline treatment in our models which might be confounding our results considering some studies have shown that MS disability accrual rates are lessening over time and MS neurologists are increasingly using higher efficacy DMTs earlier.

One of the limitations of propensity adjustment is that it only adjusts for measured, known factors, whereas other factors, unmeasured or unknown, may affect clinicians' prescribing decision and may subsequently affect what type of patient is treated with what therapy. These unmeasured or unknown factors can introduce a hidden bias which are not accounted for and can impact on the results. This is why RCTs are still needed and remain to be the gold standard for evaluating and comparing effectiveness of multiple therapies.

In our data gathering and analysis, we did not include spinal cord MRI findings that could detect further MS activity, particularly for patients with predominantly spinal cord disease, due to lack of data. (264) This is because spinal cord imaging is not routinely performed in clinical practice in the NHS particularly at follow ups, unless indicated, to assess treatment response. (68) Spinal cord imaging analysis remains difficult due to the lack of standardised imaging techniques over the past two decades and the difficulty of obtaining volumetric imaging in a very small area which is liable to motion artefact. (265) Having said that, in our ongoing prospective PITMS study, our recruited patients are having spinal cord imaging, according to the 2021 MAGNIMS-CMSC-NAIMS consensus recommendations, at each time-point of assessment, which will potentially guide spinal cord imaging standardisation and evaluation.

We evaluated the effectiveness of therapies over a two-year period which could potentially obscure long-term disease activity. Studies have shown that stability of MS may take several years and early treatment reduces the risk of MS activity, particularly relapse-associated worsening (RAW). However, MS may progress over time with ongoing disease progression independent of relapse activity (PIRA), which becomes the primary driver in the development of SPMS. (266) This is why we have set up a database that can be updated regularly and future analysis over a longer period of follow up can be performed.

A small proportion of our patients in our cohort had EDSS score of >6.5 at baseline, which is outside NHS England prescribing guidelines for commencing DMTs. These patients received therapy following multidisciplinary team discussions as they were deemed appropriate for these selected cases. However, it is well known that higher EDSS is associated with worse outcomes in RRMS patients and greater chance of developing SPMS. Another limitation to add is that our study was conducted at a single tertiary MS centre, both of which can possibly limit generalisability to other centres as clinicians can differ in their practice of counselling and prescribing different DMTs. On the other hand, the monocentric nature of our study possibly provides standardised clinical assessment and management as patients commence therapy

following multidisciplinary team discussions, particularly for higher efficacy therapies.

In our study we used the absolute change in EDSS from baseline to year one and two and we have not assessed the proportion free from sustained accumulation of disability events. It is therefore not known if any EDSS changes were sustained and the information about time to event if lost. This is expected considering the data included assessment of patients on annual basis as per clinical practice.

In conclusion, we provide important class III information for evaluating the effectiveness of different DMTs in the management of RRMS. Highly effective DMTs are associated with better outcomes compared to moderately effective DMTs in terms of real-world effectiveness. Patients taking moderately effective DMTs were at higher risk of experiencing disease activity in the first two years. Randomised controlled trials and further long-term studies evaluating more DMTs in larger cohorts, so that more robust statistics such as matching analysis can be used, are needed to investigate risk factors for patients experiencing disease activity.

Chapter 6. Conclusion and future direction

6.1 Conclusion

In this thesis we initially reviewed the causes, pathology and pathophysiology of MS. We explored the use of MRI scans in the diagnosis and treatment monitoring of MS. We also explored the different therapy strategies including symptoms management and disease modifying therapies available for the holistic management of MS.

Due to the heterogeneity of onset, diagnosis, classification, treatment response and prognosis, it is difficult to predict outcomes in MS patients. Therefore, in chapter two, We reviewed the literature on the use of machine learning in MS diagnosis, classification, progression and prediction. Studies have focused on predicting disability progression and the response to the large number of available therapies in order to help clinicians to choose the best treatment for MS patients. Despite the rapidly evolving field of ML in MS, there have been a limited number of studies focusing on the prediction of treatment responses.

In chapter three, we reviewed and analysed clinical and radiological data of nearly 2000 RRMS patients who had initiated five DMTs at a single tertiary MS centre over the last two decades. We investigated the impact of each variable, observed in each DMT cohort separately, on the outcome, NEDA, using the frequentist analysis. Reviewing the baseline variables across all cohorts; older age at MS onset, lower number of previous DMTs, lower number of relapses in the previous 12 months, lower number of new and/or Gd-enhancing MRI brain lesions, and lower EDSS score showed significant association of achieving NEDA in most cohorts. Whilst these factors are associated with clinical response at the group-level, we were unable to make clinically useful predictions of treatment response at the individual-level.

In chapter four, we built a Bayesian modelling framework optimal for estimating individual-level uncertainty in predicting NEDA in RRMS patients starting DMTs using routinely available clinical and radiological data. We combined all cohorts and built five predicting models with increasing complexity. The most comprehensive model

with 12 clinical and radiological predictors showed the best accuracy in predicting individualised NEDA at 1 and 2 years.

Finally, there has been no head-to-head phase III clinical trials comparing the efficacy of multiple DMTs in RRMS. Therefore, in chapter five, we attempted to compare the clinical and radiological effectiveness of five different DMTs in our real-world cohort. However, there were several limitations to our study, including using propensity score adjusting rather than matching, and further long-term studies evaluating and comparing more DMTs in larger cohorts are needed to investigate risk factors for patients experiencing disease activity.

Although our work showed results that are important in the MS field, it is also important to demonstrate the limitations. The cohort analysed is obtained from a single tertiary centre, representing a particular population and therefore interpretation in other populations should be cautious for the time being, until validation in an unseen independent cohort is finalised.

The data collected was mixed and the majority was collected retrospectively and a small proportion, ocrelizumab patients, was collected prospectively as part of the on-going prospective observational PITMS study. The DMT cohorts were collected over many years, the criteria for prescribing DMTs has changed and evolved over this time, therefore this may affect the level of DMTs used for the level of MS activity observed. The MRI data was collected from scans of different time periods with different protocols that have evolved over time. This has made MRI analysis challenging particularly in the hope of developing a fully automated pipeline.

Finally, our analysis included a short period of follow up. Considering our cohort had an average duration of being on therapy of nearly 4 years, further analysis of longer-term outcomes is planned.

6.2 Future direction

Study Design

Our work has demonstrated a large cohort of RRMS patients analysed from a single tertiary MS centre over a long period of time. The use of DMTs has changed over time and other therapies, which are not included in this thesis, are also being routinely prescribed. These including ofatumumab, ponesimod and clardribine; data of RRMS patient cohorts taking these other DMTs are prospectively being evaluated. This will generate a further sizable cohort that can be combined and assessed alongside the cohort included in this study for a more powerful analysis, particularly when comparing effectiveness (for example in chapter five) and the possibility of using propensity score matching. In the meantime, we are in the process of validating the Bayesian models in an unseen, independent Swedish RRMS cohort. In addition, our department is extending the investigation of prediction of treatment response to other NHS trusts in the UK by 'decentralising' artificial intelligence and using federated methods. Decentralisation will involve raw data being kept locally within each NHS trust, and the insights generated from local analytics being sent to a server for result aggregation. This is an important step to ensure a larger number of participants (up to 15,000 MS patients on DMTs) and develop a 'decision aid system', which will help to improve patient care and treatment.

More complex variables, including serum biomarkers (such as serum neurofilament), diet, genetics and optical coherence tomography (OCT) measurement are being explored through the on-going prospective PITMS study. Although these variables have been studied separately, they have not been assessed in combination with many clinical and radiological variables to predict the short- and long-term treatment outcomes such as NEDA. Collaborating and accessing larger international registries, such as the MSBase, and testing our models would be of extreme importance for the future in comparing different DMTs.

There have been a number of studies looking at the two ways of commencing therapy for RRMS; escalation (starting a lower efficacy therapy and escalating as needed)

versus induction (commencing a high efficacy therapy at the start). Escalation is the more widely used treatment strategy, as induction is restricted to patients at risk of rapid disability accrual. Over time, there has been growing evidence for more favourable, long-term outcomes following early intensive therapy versus first-line moderate-efficacy DMT. (119, 267) Therefore, future observational studies are likely to include a higher number of high-efficacy drugs. So far, there has been little evidence about the optimal sequence of DMTs for the escalation strategy, and extending the length of follow-up in our cohort has the potential to capture the impact of different combinations of DMTs over time.

MS phenotypes:

Our work has been confined to RRMS patients and future work will include other types of MS, including PPMS and SPMS who commence DMTs, with longer follow up periods. This will provide further insight of treatment response and disease progression while on therapy. An important aim which will be achieved by future studies is to predict treatment response in children with MS. Paediatric MS accounts for 3-10% of all patient with MS. Current treatment regimes in paediatric MS are largely centre-specific and based predominantly on adult protocols. Children with MS requiring long-term immunosuppression are currently at risk of being either under-treated (which results in acquired motor, visual and cognitive disabilities) or over-treater, particularly with concurrent or sequential immunotherapy. This Immunotherapy has potential short- and long-term toxicities, including impact on subsequent fertility and risks of malignancy and premature immune senescence. Therefore, the prediction of treatment response in children with MS is an important unmet need.

Statistical analysis

Randomised controlled trials (RCTs) are the primary design for evaluating the difference in treatment effect between patient groups treated with different (or no) DMTs. However, RCTs are performed under optimal circumstances (e.g. over-representation of treatment-adherent patients, absence of co-morbidities, low

representation of old patients), which may be different from real-life practice. Observational studies have the advantage of limiting the issue of external validity, but patients treated with different medications (or patients not treated) are often non-comparable, as seen in our study, thereby leading to a high risk of confounding bias. Propensity score-based methods are some of the methods currently employed to mitigate the effects of confounding bias. A future, more statistical, direction of research is to explore methods other than propensity score-based methods to control for confounding bias, including doubly robust estimators and g-computation. Recent evidence suggests that considering all the covariates influencing the outcome may lead to the lowest bias and variance when using causal inference methods, particularly for g-computation. (268)

Predictive variables

More baseline variables are being explored through the on-going prospective PITMS study. These include spinal cord MRI. In particular, number of spinal cord lesions and spinal cord cross-sectional area (which is used as a proxy for atrophy) would be of importance considering that spinal cord pathology tends to be symptomatic and associated with worse outcomes over time.

Other variables which may predict treatment response include serum biomarkers (such as serum neurofilament), and optical coherence tomography (OCT) measurement. Although these variables have been studied separately, they have not been assessed in combination with many clinical and radiological variables to predict the short- and long-term treatment outcomes, such as NEDA. Variables that are more explorative and have a less significant impact on treatment response (compared to MRI, EDSS and age/sex) include diet, environmental and lifestyle, and smoking. It is expected that comorbidities can influence the effectiveness and tolerability of DMTs; for example, fingolimod-associated macular oedema is higher in patients with diabetes. (269) Genetic markers that predict response to the oldest DMTs have been detected. (270) Additionally, HLA-DRB1*15 influences the progression of brain pathology in progressive MS and is associated with a more aggressive MS course in relapsing-onset MS. (271, 272) This suggests that genetic factors may identify

patients who are at greater risk of accumulation of brain pathology, and therefore, worse outcome.

Outcome measure:

Considering we only included treatment response in the form of NEDA at 1 and 2 years, longer term response at 3 and 5 years would capture a better real-world efficacy of therapies over a longer period of time. NEDA-3 is the only outcome measure we focused on in this thesis and therefore future analysis with the inclusion of brain volume atrophy will analyse NEDA-4, assessing treatment response in terms of relapses, MRI activity, EDSS worsening and brain atrophy rate. NEDA can be criticised for being too strict as patients would fail NEDA if they develop one new asymptomatic brain lesion or one significant spinal cord relapse. Future analysis of minimal disease activity (MEDA) can be explored in the short and long term.

In addition, we tend to estimate treatment response by studying disease activity or progression, without focusing on side effects. Future studies will investigate the prediction of serious side effects that lead to treatment discontinuation. In our PITMS study we have collected data including the side effects experienced by all patients; therefore, it would be interesting to test our models in predicting the risk of side effects in the short- and long-term for the individual patient. This would help the clinicians and patients deciding on the best treatment not just in terms of efficacy, but also in terms of tolerance.

Cognition, fatigue and quality of life measures are important to assess for MS patients. Although it is difficult to include in retrospective NHS patients where such data is not collected routinely in the clinical setting, future prospective study (such as our PITMS study) will collect such data, including the Brief International Cognitive Assessment for MS (BICAMS) battery, which includes tests of mental processing speed and memory. (273) One of the tests of the BICAMS is the symbol digit modalities test (SDMT). (274) Ubiquitously employed in cross-sectional research and clinical trials, the SDMT has high validity for discriminating between people with and without MS, and predicts real-world functional outcomes such as employment status.

(275) However, for detecting cognitive decline over time, the SDMT is inadequate, as inspection of change on SDMT over time reveals overall improvement, which is due to vulnerability of SDMT to practice effects. (276) Therefore, future studies will aim to develop and employ cognitive measurement tools sensitive to subtle decline over time as outcome measures.

Patient reported outcomes, including MS quality of life and modified fatigue impact scales, will be explored as a measure of outcomes in future analysis. Important outcome measures, which are rarely included in observational studies, are health-economics scales and information related to employment, education and absenteeism, as these measures also capture the effect of DMTs.

Integration into routine clinical workflows:

Since patients' details in the NHS are being increasingly stored in the electronic health record, it would be of a huge benefit to embed the phenotypic collection of clinical and paraclinical variables into patients' clinical care. This would make access and analysis of records much easier and quicker. Also, the ultimate aim, once our models are finalised and validated, is to develop as a decision aid system that generates prediction in clinic, such as the Fracture Risk Assessment Tool (FRAX) that is used for calculating the 10-year risk of a fracture in patients with osteoporosis, to provide the best treatment for the individual patient.

Once a decision aid system is created, it is important to investigate whether it can improve practice and clinical care. A pragmatic randomised clinical trial was recently reported using artificial intelligence to diagnose low ejection fraction through electrocardiograms. (277) Similar pragmatic trial using MRI brain scans to diagnose MS early can be conducted in the future in MS

Chapter 7. References

1. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet. 2018;391(10130):1622-36.
2. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler. 2020;26(14):1816-21.
3. Mackenzie IS, Morant SV, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990-2010: a descriptive study in the General Practice Research Database. J Neurol Neurosurg Psychiatry. 2014;85(1):76-84.
4. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology. 1996;46(4):907-11.
5. Klineova S, Lublin FD. Clinical Course of Multiple Sclerosis. Cold Spring Harb Perspect Med. 2018;8(9).
6. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372(9648):1502-17.
7. Brownlee WJ, Altmann DR, Prados F, Miszkiel KA, Eshaghi A, Gandini Wheeler-Kingshott CAM, et al. Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. Brain. 2019;142(8):2276-87.
8. Chung KK, Altmann D, Barkhof F, Miszkiel K, Brex PA, O'Riordan J, et al. A 30-Year Clinical and Magnetic Resonance Imaging Observational Study of Multiple Sclerosis and Clinically Isolated Syndromes. Ann Neurol. 2020;87(1):63-74.
9. Fambiatos A, Jokubaitis V, Horakova D, Kubala Havrdova E, Trojano M, Prat A, et al. Risk of secondary progressive multiple sclerosis: A longitudinal study. Mult Scler. 2020;26(1):79-90.
10. Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J, et al. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. Brain. 1989;112 (Pt 6):1419-28.
11. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. N Engl J Med. 2000;343(20):1430-8.
12. Kremenchutzky M, Rice GP, Baskerville J, Wingerchuk DM, Ebers GC. The natural history of multiple sclerosis: a geographically based study 9: observations on the progressive phase of the disease. Brain. 2006;129(Pt 3):584-94.
13. Cree BAC, Arnold DL, Chataway J, Chitnis T, Fox RJ, Pozo Ramajo A, et al. Secondary Progressive Multiple Sclerosis: New Insights. Neurology. 2021;97(8):378-88.
14. Rovaris M, Confavreux C, Furlan R, Kappos L, Comi G, Filippi M. Secondary progressive multiple sclerosis: current knowledge and future challenges. Lancet Neurol. 2006;5(4):343-54.
15. University of California SFMSET, Cree BA, Gourraud PA, Oksenberg JR, Bevan C, Crabtree-Hartman E, et al. Long-term evolution of multiple sclerosis disability in the treatment era. Ann Neurol. 2016;80(4):499-510.
16. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-86.

17. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science. 2022;375(6578):296-301.
18. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA. 2006;296(23):2832-8.
19. Sintzel MB, Rametta M, Reder AT. Vitamin D and Multiple Sclerosis: A Comprehensive Review. Neurol Ther. 2018;7(1):59-85.
20. Hedstrom AK, Olsson T, Alfredsson L. Smoking is a major preventable risk factor for multiple sclerosis. Mult Scler. 2016;22(8):1021-6.
21. Rosso M, Chitnis T. Association Between Cigarette Smoking and Multiple Sclerosis: A Review. JAMA Neurol. 2020;77(2):245-53.
22. Simpson S, Jr., Wang W, Otahal P, Blizzard L, van der Mei IAF, Taylor BV. Latitude continues to be significantly associated with the prevalence of multiple sclerosis: an updated meta-analysis. J Neurol Neurosurg Psychiatry. 2019;90(11):1193-200.
23. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, Canadian Collaborative Study G. Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci U S A. 2003;100(22):12877-82.
24. Kantarci OH. Genetics and natural history of multiple sclerosis. Semin Neurol. 2008;28(1):7-16.
25. Compston DA, Batchelor JR, McDonald WI. B-lymphocyte alloantigens associated with multiple sclerosis. Lancet. 1976;2(7998):1261-5.
26. Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens. 1991;38(1):1-15.
27. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
28. Hoppenbrouwers IA, Hintzen RQ. Genetics of multiple sclerosis. Biochim Biophys Acta. 2011;1812(2):194-201.
29. Zajicek JP, Wing M, Scolding NJ, Compston DA. Interactions between oligodendrocytes and microglia. A major role for complement and tumour necrosis factor in oligodendrocyte adherence and killing. Brain. 1992;115 (Pt 6):1611-31.
30. Frischer JM, Weigand SD, Guo Y, Kale N, Parisi JE, Pirko I, et al. Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque. Ann Neurol. 2015;78(5):710-21.
31. Kutzelnigg A, Lucchinetti CF, Stadelmann C, Brück W, Rauschka H, Bergmann M, et al. Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain. 2005;128(Pt 11):2705-12.
32. Kebir H, Kreymborg K, Ifergan I, Dodelet-Devillers A, Cayrol R, Bernard M, et al. Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. Nat Med. 2007;13(10):1173-5.
33. Tzartos JS, Friese MA, Craner MJ, Palace J, Newcombe J, Esiri MM, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. Am J Pathol. 2008;172(1):146-55.

34. Kuhlmann T, Lingfeld G, Bitsch A, Schuchardt J, Brück W. Acute axonal damage in multiple sclerosis is most extensive in early disease stages and decreases over time. Brain. 2002;125(Pt 10):2202-12.
35. Calabrese M, Poretto V, Favaretto A, Alessio S, Bernardi V, Romualdi C, et al. Cortical lesion load associates with progression of disability in multiple sclerosis. Brain. 2012;135(Pt 10):2952-61.
36. Filippi M, Preziosa P, Pagani E, Copetti M, Mesaros S, Colombo B, et al. Microstructural magnetic resonance imaging of cortical lesions in multiple sclerosis. Mult Scler. 2013;19(4):418-26.
37. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mork S, Bo L. Axonal transection in the lesions of multiple sclerosis. N Engl J Med. 1998;338(5):278-85.
38. Bitsch A, Schuchardt J, Bunkowski S, Kuhlmann T, Bruck W. Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation. Brain. 2000;123 (Pt 6):1174-83.
39. Youl BD, Turano G, Miller DH, Towell AD, MacManus DG, Moore SG, et al. The pathophysiology of acute optic neuritis. An association of gadolinium leakage with clinical and electrophysiological deficits. Brain. 1991;114 (Pt 6):2437-50.
40. Smith KJ. Sodium channels and multiple sclerosis: roles in symptom production, damage and therapy. Brain Pathol. 2007;17(2):230-42.
41. Miller DH, Rudge P, Johnson G, Kendall BE, Macmanus DG, Moseley IF, et al. Serial gadolinium enhanced magnetic resonance imaging in multiple sclerosis. Brain. 1988;111 (Pt 4):927-39.
42. Craner MJ, Hains BC, Lo AC, Black JA, Waxman SG. Co-localization of sodium channel Nav1.6 and the sodium-calcium exchanger at sites of axonal injury in the spinal cord in EAE. Brain. 2004;127(Pt 2):294-303.
43. Patani R, Balaratnam M, Vora A, Reynolds R. Remyelination can be extensive in multiple sclerosis despite a long disease course. Neuropathol Appl Neurobiol. 2007;33(3):277-87.
44. Toosy AT, Hickman SJ, Miszkiel KA, Jones SJ, Plant GT, Altmann DR, et al. Adaptive cortical plasticity in higher visual areas after acute optic neuritis. Ann Neurol. 2005;57(5):622-33.
45. Smith KJ, McDonald WI. Spontaneous and evoked electrical discharges from a central demyelinating lesion. J Neurol Sci. 1982;55(1):39-47.
46. Lhermitte J. Les formes douloureuses de la commotion de la moelle épinière. Rev Neurol. 1920;36:257-62.
47. Uhthoff W. Untersuchungen über die bei der multiplen Herdsklerose vorkommenden Augenstörungen. Archiv für Psychiatrie und Nervenkrankheiten 1890(21):55-116.
48. Bostock H, Grafe P. Activity-dependent excitability changes in normal and demyelinated rat spinal root axons. J Physiol. 1985;365:239-57.
49. Li DK, Held U, Petkau J, Daumer M, Barkhof F, Fazekas F, et al. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. Neurology. 2006;66(9):1384-9.
50. Evangelou N, DeLuca GC, Owens T, Esiri MM. Pathological study of spinal cord atrophy in multiple sclerosis suggests limited role of local lesions. Brain. 2005;128(Pt 1):29-34.

51. Miller DH, Barkhof F, Frank JA, Parker GJ, Thompson AJ. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. Brain. 2002;125(Pt 8):1676-95.
52. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33(11):1444-52.
53. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain. 1999;122 (Pt 5):871-82.
54. Brichetto G, Zaratin P. Measuring outcomes that matter most to people with multiple sclerosis: the role of patient-reported outcomes. Curr Opin Neurol. 2020;33(3):295-9.
55. Essig M, Nguyen TB, Shiroishi MS, Saake M, Provenzale JM, Enterline DS, et al. Perfusion MRI: the five most frequently asked clinical questions. AJR Am J Roentgenol. 2013;201(3):W495-510.
56. van Walderveen MA, Kamphorst W, Scheltens P, van Waesberghe JH, Ravid R, Valk J, et al. Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. Neurology. 1998;50(5):1282-8.
57. Truyen L, van Waesberghe JH, van Walderveen MA, van Oosten BW, Polman CH, Hommes OR, et al. Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. Neurology. 1996;47(6):1469-76.
58. Stevenson VL, Miller DH. Magnetic resonance imaging in the monitoring of disease progression in multiple sclerosis. Mult Scler. 1999;5(4):268-72.
59. Thompson AJ, Montalban X, Barkhof F, Brochet B, Filippi M, Miller DH, et al. Diagnostic criteria for primary progressive multiple sclerosis: a position paper. Ann Neurol. 2000;47(6):831-5.
60. Barkhof F, Bruck W, De Groot CJ, Bergers E, Hulshof S, Geurts J, et al. Remyelinated lesions in multiple sclerosis: magnetic resonance image appearance. Arch Neurol. 2003;60(8):1073-81.
61. Geurts JJ, Bo L, Pouwels PJ, Castelijns JA, Polman CH, Barkhof F. Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. AJNR Am J Neuroradiol. 2005;26(3):572-7.
62. Hashemi RH, Bradley WG, Chen DY, Jordan JE, Queralt JA, Cheng AE, et al. Suspected multiple sclerosis: MR imaging with a thin-section fast FLAIR pulse sequence. Radiology. 1995;196(2):505-10.
63. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292-302.
64. Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? Mult Scler Relat Disord. 2015;4(4):329-33.
65. Honce JM, Nagae L, Nyberg E. Neuroimaging of Natalizumab Complications in Multiple Sclerosis: PML and Other Associated Entities. Mult Scler Int. 2015;2015:809252.
66. Christoforidis GA, Spickler EM, Recio MV, Mehta BM. MR of CNS sarcoidosis: correlation of imaging features to clinical symptoms and response to treatment. AJNR Am J Neuroradiol. 1999;20(4):655-69.

67. Sammet S. Magnetic resonance safety. Abdom Radiol (NY). 2016;41(3):444-51.
68. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol. 2021;20(8):653-70.
69. Anderson VM, Fox NC, Miller DH. Magnetic resonance imaging measures of brain atrophy in multiple sclerosis. J Magn Reson Imaging. 2006;23(5):605-18.
70. Dalton CM, Chard DT, Davies GR, Miszkiel KA, Altmann DR, Fernando K, et al. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain. 2004;127(Pt 5):1101-7.
71. Eshaghi A, Prados F, Brownlee WJ, Altmann DR, Tur C, Cardoso MJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol. 2018;83(2):210-22.
72. Ingle GT, Stevenson VL, Miller DH, Leary SM, Rovaris M, Barkhof F, et al. Two-year follow-up study of primary and transitional progressive multiple sclerosis. Mult Scler. 2002;8(2):108-14.
73. Gasperini C, Paolillo A, Giugni E, Galgani S, Bagnato F, Mainero C, et al. MRI brain volume changes in relapsing-remitting multiple sclerosis patients treated with interferon beta-1a. Mult Scler. 2002;8(2):119-23.
74. Fisher E, Rudick RA, Simon JH, Cutter G, Baier M, Lee JC, et al. Eight-year follow-up study of brain atrophy in patients with MS. Neurology. 2002;59(9):1412-20.
75. Chard DT, Griffin CM, Rashid W, Davies GR, Altmann DR, Kapoor R, et al. Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. Mult Scler. 2004;10(4):387-91.
76. Sastre-Garriga J, Ingle GT, Chard DT, Ramio-Torrenta L, Miller DH, Thompson AJ. Grey and white matter atrophy in early clinical stages of primary progressive multiple sclerosis. Neuroimage. 2004;22(1):353-9.
77. Furby J, Hayton T, Anderson V, Altmann D, Brenner R, Chataway J, et al. Magnetic resonance imaging measures of brain and spinal cord atrophy correlate with clinical impairment in secondary progressive multiple sclerosis. Mult Scler. 2008;14(8):1068-75.
78. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121-7.
79. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol. 2005;58(6):840-6.
80. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162-73.
81. Gobbin F, Zanoni M, Marangi A, Orlandi R, Crestani L, Benedetti MD, et al. 2017 McDonald criteria for multiple sclerosis: Earlier diagnosis with reduced specificity? Mult Scler Relat Disord. 2019;29:23-5.

82. van der Vuurst de Vries RM, Mescheriakova JY, Wong YYM, Runia TF, Jafari N, Samijn JP, et al. Application of the 2017 Revised McDonald Criteria for Multiple Sclerosis to Patients With a Typical Clinically Isolated Syndrome. JAMA Neurol. 2018;75(11):1392-8.
83. Habek M, Pavicic T, Ruska B, Pavlovic I, Gabelic T, Barun B, et al. Establishing the diagnosis of multiple sclerosis in Croatian patients with clinically isolated syndrome: 2010 versus 2017 McDonald criteria. Mult Scler Relat Disord. 2018;25:99-103.
84. Hyun JW, Kim W, Huh SY, Park MS, Ahn SW, Cho JY, et al. Application of the 2017 McDonald diagnostic criteria for multiple sclerosis in Korean patients with clinically isolated syndrome. Mult Scler. 2019;25(11):1488-95.
85. Lee DH, Peschke M, Utz KS, Linker RA. Diagnostic value of the 2017 McDonald criteria in patients with a first demyelinating event suggestive of relapsing-remitting multiple sclerosis. Eur J Neurol. 2019;26(3):540-5.
86. Lorscheider J, Buzzard K, Jokubaitis V, Spelman T, Havrdova E, Horakova D, et al. Defining secondary progressive multiple sclerosis. Brain. 2016;139(Pt 9):2395-405.
87. Le Page E, Veillard D, Laplaud DA, Hamonic S, Wardi R, Lebrun C, et al. Oral versus intravenous high-dose methylprednisolone for treatment of relapses in patients with multiple sclerosis (COPOUSEP): a randomised, controlled, double-blind, non-inferiority trial. Lancet. 2015;386(9997):974-81.
88. Ramo-Tello C, Grau-Lopez L, Tintore M, Rovira A, Ramio i Torrenta L, Brieva L, et al. A randomized clinical trial of oral versus intravenous methylprednisolone for relapse of MS. Mult Scler. 2014;20(6):717-25.
89. Multiple Sclerosis: Management of Multiple Sclerosis in Primary and Secondary Care. National Institute for Health and Care Excellence: Clinical Guidelines. London2014.
90. Rolfes L, Pfeuffer S, Ruck T, Melzer N, Pawlitzki M, Heming M, et al. Therapeutic Apheresis in Acute Relapsing Multiple Sclerosis: Current Evidence and Unmet Needs-A Systematic Review. J Clin Med. 2019;8(10).
91. McGinley MP, Goldschmidt CH, Rae-Grant AD. Diagnosis and Treatment of Multiple Sclerosis: A Review. JAMA. 2021;325(8):765-79.
92. Bakshi R. Fatigue associated with multiple sclerosis: diagnosis, impact and management. Mult Scler. 2003;9(3):219-27.
93. Fox RJ, Bacon TE, Chamot E, Salter AR, Cutter GR, Kalina JT, et al. Prevalence of multiple sclerosis symptoms across lifespan: data from the NARCOMS Registry. Neurodegener Dis Manag. 2015;5(6 Suppl):3-10.
94. Tur C. Fatigue Management in Multiple Sclerosis. Curr Treat Options Neurol. 2016;18(6):26.
95. Shi J, Wu X, Chen Y. Study on Dalfampridine in the treatment of Multiple Sclerosis Mobility Disability: A meta-analysis. PLoS One. 2019;14(9):e0222288.
96. Litwiller SE, Frohman EM, Zimmern PE. Multiple sclerosis and the urologist. J Urol. 1999;161(3):743-57.
97. Aharony SM, Lam O, Corcos J. Evaluation of lower urinary tract symptoms in multiple sclerosis patients: Review of the literature and current guidelines. Can Urol Assoc J. 2017;11(1-2):61-4.

98. Rizzo MA, Hadjimichael OC, Preiningerova J, Vollmer TL. Prevalence and treatment of spasticity reported by multiple sclerosis patients. Mult Scler. 2004;10(5):589-95.
99. Thompson AJ, Jarrett L, Lockley L, Marsden J, Stevenson VL. Clinical management of spasticity. J Neurol Neurosurg Psychiatry. 2005;76(4):459-63.
100. Hinds JP, Eidelman BH, Wald A. Prevalence of bowel dysfunction in multiple sclerosis. A population survey. Gastroenterology. 1990;98(6):1538-42.
101. Preziosi G, Gordon-Dixon A, Emmanuel A. Neurogenic bowel dysfunction in patients with multiple sclerosis: prevalence, impact, and management strategies. Degener Neurol Neuromuscul Dis. 2018;8:79-90.
102. Martinelli Boneschi F, Colombo B, Annovazzi P, Martinelli V, Bernasconi L, Solaro C, et al. Lifetime and actual prevalence of pain and headache in multiple sclerosis. Mult Scler. 2008;14(4):514-21.
103. Murphy KL, Bethea JR, Fischer R. Neuropathic Pain in Multiple Sclerosis-Current Therapeutic Intervention and Future Treatment Perspectives. In: Zagon IS, McLaughlin PJ, editors. Multiple Sclerosis: Perspectives in Treatment and Pathogenesis. Brisbane (AU)2017.
104. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. The IFNB Multiple Sclerosis Study Group. Neurology. 1993;43(4):655-61.
105. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. Lancet. 1998;352(9139):1498-504.
106. Calabresi PA, Kieseier BC, Arnold DL, Balcer LJ, Boyko A, Pelletier J, et al. Pegylated interferon beta-1a for relapsing-remitting multiple sclerosis (ADVANCE): a randomised, phase 3, double-blind study. Lancet Neurol. 2014;13(7):657-65.
107. Jacobs LD, Beck RW, Simon JH, Kinkel RP, Brownscheidle CM, Murray TJ, et al. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group. N Engl J Med. 2000;343(13):898-904.
108. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. Neurology. 1995;45(7):1268-76.
109. O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. N Engl J Med. 2011;365(14):1293-303.
110. Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. N Engl J Med. 2012;367(12):1087-97.
111. Kappos L, Radue EW, O'Connor P, Polman C, Hohlfeld R, Calabresi P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. N Engl J Med. 2010;362(5):387-401.

112. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. N Engl J Med. 2006;354(9):899-910.
113. Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung HP, et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. Lancet. 2012;380(9856):1819-28.
114. Giovannoni G, Comi G, Cook S, Rammohan K, Rieckmann P, Soelberg Sorensen P, et al. A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. N Engl J Med. 2010;362(5):416-26.
115. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. N Engl J Med. 2017;376(3):221-34.
116. Hauser SL, Bar-Or A, Cohen JA, Comi G, Correale J, Coyle PK, et al. Ofatumumab versus Teriflunomide in Multiple Sclerosis. N Engl J Med. 2020;383(6):546-57.
117. Kappos L, Fox RJ, Burcklen M, Freedman MS, Havrdova EK, Hennessy B, et al. Ponesimod Compared With Teriflunomide in Patients With Relapsing Multiple Sclerosis in the Active-Comparator Phase 3 OPTIMUM Study: A Randomized Clinical Trial. JAMA Neurol. 2021;78(5):558-67.
118. Kavaliunas A, Manouchehrinia A, Stawiarz L, Ramanujam R, Agholme J, Hedstrom AK, et al. Importance of early treatment initiation in the clinical course of multiple sclerosis. Mult Scler. 2017;23(9):1233-40.
119. He A, Merkel B, Brown JWL, Zhovits Ryerson L, Kister I, Malpas CB, et al. Timing of high-efficacy therapy for multiple sclerosis: a retrospective observational cohort study. Lancet Neurol. 2020;19(4):307-16.
120. Brown JWL, Coles A, Horakova D, Havrdova E, Izquierdo G, Prat A, et al. Association of Initial Disease-Modifying Therapy With Later Conversion to Secondary Progressive Multiple Sclerosis. JAMA. 2019;321(2):175-87.
121. Ontaneda D, Tallantyre E, Kalincik T, Planchon SM, Evangelou N. Early highly effective versus escalation treatment approaches in relapsing multiple sclerosis. Lancet Neurol. 2019;18(10):973-80.
122. Vermersch P, Suchet L, Colamarino R, Laurendeau C, Detournay B. An analysis of first-line disease-modifying therapies in patients with relapsing-remitting multiple sclerosis using the French nationwide health claims database from 2014-2017. Mult Scler Relat Disord. 2020;46:102521.
123. Sacca F, Lanzillo R, Signori A, Maniscalco GT, Signoriello E, Lo Fermo S, et al. Determinants of therapy switch in multiple sclerosis treatment-naive patients: A real-life study. Mult Scler. 2019;25(9):1263-72.
124. Maurer M, Tiel-Wilck K, Oehm E, Richter N, Springer M, Oschmann P, et al. Reasons to switch: a noninterventional study evaluating immunotherapy switches in a large German multicentre cohort of patients with relapsing-remitting multiple sclerosis. Ther Adv Neurol Disord. 2019;12:1756286419892077.
125. Tsivgoulis G, Katsanos AH, Grigoriadis N, Hadjigeorgiou GM, Heliopoulos I, Kilidireas C, et al. The effect of disease modifying therapies on brain atrophy in patients with relapsing-remitting multiple sclerosis: a systematic review and meta-analysis. PLoS One. 2015;10(3):e0116511.

126. Kappos L, Wolinsky JS, Giovannoni G, Arnold DL, Wang Q, Bernasconi C, et al. Contribution of Relapse-Independent Progression vs Relapse-Associated Worsening to Overall Confirmed Disability Accumulation in Typical Relapsing Multiple Sclerosis in a Pooled Analysis of 2 Randomized Clinical Trials. JAMA Neurol. 2020;77(9):1132-40.
127. Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ, et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. Lancet. 2012;380(9856):1829-39.
128. Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. N Engl J Med. 2017;376(3):209-20.
129. Auguste P, Colquitt J, Connock M, Loveman E, Court R, Ciccarelli O, et al. Ocrelizumab for Treating Patients with Primary Progressive Multiple Sclerosis: An Evidence Review Group Perspective of a NICE Single Technology Appraisal. Pharmacoeconomics. 2020;38(6):527-36.
130. Kappos L, Bar-Or A, Cree BAC, Fox RJ, Giovannoni G, Gold R, et al. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. Lancet. 2018;391(10127):1263-73.
131. Wise J. Multiple sclerosis: NICE recommends treatment for secondary progressive MS. BMJ. 2020;371:m3990.
132. Comi G, Radaelli M, Soelberg Sorensen P. Evolving concepts in the treatment of relapsing multiple sclerosis. Lancet. 2017;389(10076):1347-56.
133. Langer-Gould A, Popat RA, Huang SM, Cobb K, Fontoura P, Gould MK, et al. Clinical and demographic predictors of long-term disability in patients with relapsing-remitting multiple sclerosis: a systematic review. Arch Neurol. 2006;63(12):1686-91.
134. Confavreux C, Vukusic S. Age at disability milestones in multiple sclerosis. Brain. 2006;129(Pt 3):595-605.
135. Debouverie M. Gender as a prognostic factor and its impact on the incidence of multiple sclerosis in Lorraine, France. J Neurol Sci. 2009;286(1-2):14-7.
136. Seyman E, Jones A, Guenette M, Vosoughi R, Selchen D, Amezcua L, et al. Clinical and MRI characteristics of multiple sclerosis in patients of Middle Eastern and North African ancestry residing in Ontario, Canada. Mult Scler. 2021;27(7):1027-36.
137. Marrie RA, Rudick R, Horwitz R, Cutter G, Tyry T, Campagnolo D, et al. Vascular comorbidity is associated with more rapid disability progression in multiple sclerosis. Neurology. 2010;74(13):1041-7.
138. McKay KA, Tremlett H, Fisk JD, Zhang T, Patten SB, Kastrukoff L, et al. Psychiatric comorbidity is associated with disability progression in multiple sclerosis. Neurology. 2018;90(15):e1316-e23.
139. Vasconcelos CC, Aurencao JC, Thuler LC, Camargo S, Alvarenga MP, Alvarenga RM. Prognostic factors associated with long-term disability and secondary progression in patients with Multiple Sclerosis. Mult Scler Relat Disord. 2016;8:27-34.
140. Scalfari A, Neuhaus A, Degenhardt A, Rice GP, Muraro PA, Daumer M, et al. The natural history of multiple sclerosis: a geographically based study 10: relapses and long-term disability. Brain. 2010;133(Pt 7):1914-29.

141. Alroughani RA, Akhtar S, Ahmed SF, Al-Hashel JY. Clinical predictors of disease progression in multiple sclerosis patients with relapsing onset in a nation-wide cohort. Int J Neurosci. 2015;125(11):831-7.
142. Minneboo A, Barkhof F, Polman CH, Uitdehaag BM, Knol DL, Castelijns JA. Infratentorial lesions predict long-term disability in patients with initial findings suggestive of multiple sclerosis. Arch Neurol. 2004;61(2):217-21.
143. Cordonnier C, de Seze J, Breteau G, Ferriby D, Michelin E, Stojkovic T, et al. Prospective study of patients presenting with acute partial transverse myelopathy. J Neurol. 2003;250(12):1447-52.
144. Tintore M, Rovira A, Rio J, Otero-Romero S, Arrambide G, Tur C, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. Brain. 2015;138(Pt 7):1863-74.
145. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-43.
146. Kersting K. Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines. Front Big Data. 2018;1:6.
147. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. NPJ Digit Med. 2020;3:30.
148. Auger SD, Jacobs BM, Dobson R, Marshall CR, Noyce AJ. Big data, machine learning and artificial intelligence: a neurologist's guide. Pract Neurol. 2020;21(1):4-11.
149. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform. 2014;6(1):10.
150. Jung Y, Hu J. A K-fold Averaging Cross-validation Procedure. J Nonparametr Stat. 2015;27(2):167-79.
151. Brownlee J. A Gentle Introduction to Joint, Marginal, and Conditional Probability. 2019.
152. Jr FEH. Contrasting Bayesian and Frequentist Statistical Methods 2022 [https://hbiostat.org/proj/covid19/statdesign.html#contrasting-bayesian-and-frequentist-statistical-methods.
153. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philosophical Transactions of the Royal Society of London. 1997;53:370-418.
154. Bayesian Statistics explained to Beginners in Simple English. 2016.
155. Vrenken H, Jenkinson M, Pham DL, Guttmann CRG, Pareto D, Paardekooper M, et al. Opportunities for Understanding MS Mechanisms and Progression With MRI Using Large-Scale Data Sharing and Artificial Intelligence. Neurology. 2021;97(21):989-99.
156. Weygandt M, Hackmack K, Pfuller C, Bellmann-Strobl J, Paul F, Zipp F, et al. MRI pattern recognition in multiple sclerosis normal-appearing brain areas. PLoS One. 2011;6(6):e21138.
157. Richiardi J, Gschwind M, Simioni S, Annoni JM, Greco B, Hagmann P, et al. Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. Neuroimage. 2012;62(3):2021-33.

158. Eshaghi A, Wottschel V, Cortese R, Calabrese M, Sahraian MA, Thompson AJ, et al. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. Neurology. 2016;87(23):2463-70.
159. Yoo Y, Tang LYW, Brosch T, Li DKB, Kolind S, Vavasour I, et al. Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. Neuroimage Clin. 2018;17:169-78.
160. Zurita M, Montalba C, Labbe T, Cruz JP, Dalboni da Rocha J, Tejos C, et al. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. Neuroimage Clin. 2018;20:724-30.
161. Sacca V, Sarica A, Novellino F, Barone S, Tallarico T, Filippelli E, et al. Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data. Brain Imaging Behav. 2019;13(4):1103-14.
162. Azarmi F, Miri Ashtiani SN, Shalbaf A, Behnam H, Daliri MR. Granger causality analysis in combination with directed network measures for classification of MS patients and healthy controls using task-related fMRI. Comput Biol Med. 2019;115:103495.
163. Neeb H, Schenk J. Multivariate prediction of multiple sclerosis using robust quantitative MR-based image metrics. Z Med Phys. 2019;29(3):262-71.
164. Perez Del Palomar A, Cegonino J, Montolio A, Orduna E, Vilades E, Sebastian B, et al. Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. PLoS One. 2019;14(5):e0216410.
165. Cavaliere C, Vilades E, Alonso-Rodriguez MC, Rodrigo MJ, Pablo LE, Miguel JM, et al. Computer-Aided Diagnosis of Multiple Sclerosis Using a Support Vector Machine and Optical Coherence Tomography Features. Sensors (Basel). 2019;19(23).
166. Lotsch J, Thrun M, Lerch F, Brunkhorst R, Schiffmann S, Thomas D, et al. Machine-Learned Data Structures of Lipid Marker Serum Concentrations in Multiple Sclerosis Patients Differ from Those in Healthy Subjects. Int J Mol Sci. 2017;18(6).
167. Lotsch J, Schiffmann S, Schmitz K, Brunkhorst R, Lerch F, Ferreiros N, et al. Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy. Sci Rep. 2018;8(1):14884.
168. Andersen SL, Briggs FBS, Winnike JH, Natanzon Y, Maichle S, Knagge KJ, et al. Metabolome-based signature of disease pathology in MS. Mult Scler Relat Disord. 2019;31:12-21.
169. Ostmeyer J, Christley S, Rounds WH, Toby I, Greenberg BM, Monson NL, et al. Statistical classifiers for diagnosing disease from immune repertoires: a case study using multiple sclerosis. BMC Bioinformatics. 2017;18(1):401.
170. Goyal M, Khanna D, Rana PS, Khaibullin T, Martynova E, Rizvanov AA, et al. Computational Intelligence Technique for Prediction of Multiple Sclerosis Based on Serum Cytokines. Front Neurol. 2019;10:781.
171. Ghafouri-Fard S, Taheri M, Omrani MD, Daaee A, Mohammad-Rahimi H. Application of Artificial Neural Network for Prediction of Risk of Multiple Sclerosis

Based on Single Nucleotide Polymorphism Genotypes. J Mol Neurosci. 2020;70(7):1081-7.
172. Forbes JD, Chen CY, Knox NC, Marrie RA, El-Gabalawy H, de Kievit T, et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? Microbiome. 2018;6(1):221.
173. Bang S, Yoo D, Kim SJ, Jhang S, Cho S, Kim H. Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. Sci Rep. 2019;9(1):10189.
174. Ahmadi A, Davoudi S, Daliri MR. Computer Aided Diagnosis System for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention. Comput Methods Programs Biomed. 2019;169:9-18.
175. Bendfeldt K, Kloppel S, Nichols TE, Smieskova R, Kuster P, Traud S, et al. Multivariate pattern classification of gray matter pathology in multiple sclerosis. Neuroimage. 2012;60(1):400-8.
176. Crimi A, Commowick O, Maarouf A, Ferre JC, Bannier E, Tourbah A, et al. Predictive value of imaging markers at multiple sclerosis disease onset based on gadolinium- and USPIO-enhanced MRI and machine learning. PLoS One. 2014;9(4):e93024.
177. Weygandt M, Hummel HM, Schregel K, Ritter K, Allefeld C, Dommes E, et al. MRI-based diagnostic biomarkers for early onset pediatric multiple sclerosis. Neuroimage Clin. 2015;7:400-8.
178. Kocevar G, Stamile C, Hannoun S, Cotton F, Vukusic S, Durand-Dubief F, et al. Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses. Front Neurosci. 2016;10:478.
179. Muthuraman M, Fleischer V, Kolber P, Luessi F, Zipp F, Groppa S. Structural Brain Network Characteristics Can Differentiate CIS from Early RRMS. Front Neurosci. 2016;10:14.
180. Zhong J, Chen DQ, Nantes JC, Holmes SA, Hodaie M, Koski L. Combined structural and functional patterns discriminating upper limb motor disability in multiple sclerosis using multivariate approaches. Brain Imaging Behav. 2017;11(3):754-68.
181. Ion-Margineanu A, Kocevar G, Stamile C, Sima DM, Durand-Dubief F, Van Huffel S, et al. Machine Learning Approach for Classifying Multiple Sclerosis Courses by Combining Clinical Data with Lesion Loads and Magnetic Resonance Metabolic Features. Front Neurosci. 2017;11:398.
182. Mato-Abad V, Labiano-Fontcuberta A, Rodriguez-Yanez S, Garcia-Vazquez R, Munteanu CR, Andrade-Garda J, et al. Classification of radiologically isolated syndrome and clinically isolated syndrome with machine-learning techniques. Eur J Neurol. 2019;26(7):1000-5.
183. Gonzalez Campo C, Salamone PC, Rodriguez-Arriagada N, Richter F, Herrera E, Bruno D, et al. Fatigue in multiple sclerosis is associated with multimodal interoceptive abnormalities. Mult Scler. 2020;26(14):1845-53.
184. Fiorini S, Verri A, Tacchino A, Ponzio M, Brichetto G, Barla A. A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:4443-6.

185. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform. 2018;85:30-9.
186. Corvol JC, Pelletier D, Henry RG, Caillier SJ, Wang J, Pappas D, et al. Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. Proc Natl Acad Sci U S A. 2008;105(33):11839-44.
187. Wottschel V, Alexander DC, Kwok PP, Chard DT, Stromillo ML, De Stefano N, et al. Predicting outcome in clinically isolated syndrome using machine learning. Neuroimage Clin. 2015;7:281-7.
188. Zhang H, Alberts E, Pongratz V, Muhlau M, Zimmer C, Wiestler B, et al. Predicting conversion from clinically isolated syndrome to multiple sclerosis-An imaging-based machine learning approach. Neuroimage Clin. 2019;21:101593.
189. Wottschel V, Chard DT, Enzinger C, Filippi M, Frederiksen JL, Gasperini C, et al. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. Neuroimage Clin. 2019;24:102011.
190. Seccia R, Gammelli D, Dominici F, Romano S, Landi AC, Salvetti M, et al. Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis. PLoS One. 2020;15(3):e0230219.
191. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. PLoS One. 2017;12(4):e0174866.
192. Kiiski H, Jollans L, Donnchadha SO, Nolan H, Lonergan R, Kelly S, et al. Machine Learning EEG to Predict Cognitive Functioning and Processing Speed Over a 2-Year Period in Multiple Sclerosis Patients and Controls. Brain Topogr. 2018;31(3):346-63.
193. Pitteri M, Romualdi C, Magliozzi R, Monaco S, Calabrese M. Cognitive impairment predicts disability progression and cortical thinning in MS: An 8-year study. Mult Scler. 2017;23(6):848-54.
194. Law MT, Traboulsee AL, Li DK, Carruthers RL, Freedman MS, Kolind SH, et al. Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. Mult Scler J Exp Transl Clin. 2019;5(4):2055217319885983.
195. Jackson KC, Sun K, Barbour C, Hernandez D, Kosa P, Tanigawa M, et al. Genetic model of MS severity predicts future accumulation of disability. Ann Hum Genet. 2020;84(1):1-10.
196. Yperman J, Becker T, Valkenborg D, Popescu V, Hellings N, Wijmeersch BV, et al. Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. BMC Neurol. 2020;20(1):105.
197. Roca P, Attye A, Colas L, Tucholka A, Rubini P, Cackowski S, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. Diagn Interv Imaging. 2020;101(12):795-802.
198. Baranzini SE, Madireddy LR, Cromer A, D'Antonio M, Lehr L, Beelke M, et al. Prognostic biomarkers of IFNb therapy in multiple sclerosis patients. Mult Scler. 2015;21(7):894-904.

199. Kalincik T, Manouchehrinia A, Sobisek L, Jokubaitis V, Spelman T, Horakova D, et al. Towards personalized therapy for multiple sclerosis: prediction of individual treatment response. Brain. 2017;140(9):2426-43.
200. Kanber B, Nachev P, Barkhof F, Calvi A, Cardoso J, Cortese R, et al. High-dimensional detection of imaging response to treatment in multiple sclerosis. NPJ Digit Med. 2019;2:49.
201. Fagone P, Mazzon E, Mammana S, Di Marco R, Spinasanta F, Basile MS, et al. Identification of CD4+ T cell biomarkers for predicting the response of patients with relapsingremitting multiple sclerosis to natalizumab treatment. Mol Med Rep. 2019;20(1):678-84.
202. Ebrahimkhani S, Beadnall HN, Wang C, Suter CM, Barnett MH, Buckland ME, et al. Serum Exosome MicroRNAs Predict Multiple Sclerosis Disease Activity after Fingolimod Treatment. Mol Neurobiol. 2020;57(2):1245-58.
203. Jin T, Wang C, Tian S. Feature selection based on differentially correlated gene pairs reveals the mechanism of IFN-beta therapy for multiple sclerosis. PeerJ. 2020;8:e8812.
204. Giovannoni G, Tomic D, Bright JR, Havrdova E. "No evident disease activity": The use of combined assessments in the management of patients with multiple sclerosis. Mult Scler. 2017;23(9):1179-87.
205. Banwell B, Giovannoni G, Hawkes C, Lublin F. Editors' welcome and a working definition for a multiple sclerosis cure. Mult Scler Relat Disord. 2013;2(2):65-7.
206. Kappos L, De Stefano N, Freedman MS, Cree BA, Radue EW, Sprenger T, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. Mult Scler. 2016;22(10):1297-305.
207. Kalincik T, Brown JWL, Robertson N, Willis M, Scolding N, Rice CM, et al. Treatment effectiveness of alemtuzumab compared with natalizumab, fingolimod, and interferon beta in relapsing-remitting multiple sclerosis: a cohort study. Lancet Neurol. 2017;16(4):271-81.
208. Sormani MP, Gasperini C, Romeo M, Rio J, Calabrese M, Cocco E, et al. Assessing response to interferon-beta in a multicenter dataset of patients with MS. Neurology. 2016;87(2):134-40.
209. Granella F, Tsantes E, Graziuso S, Bazzurri V, Crisi G, Curti E. Spinal cord lesions are frequently asymptomatic in relapsing-remitting multiple sclerosis: a retrospective MRI survey. J Neurol. 2019;266(12):3031-7.
210. Sormani MP, Freedman MS, Aldridge J, Marhardt K, Kappos L, De Stefano N. MAGNIMS score predicts long-term clinical disease activity-free status and confirmed disability progression in patients treated with subcutaneous interferon beta-1a. Mult Scler Relat Disord. 2021;49:102790.
211. Tintore M, Rovira A, Arrambide G, Mitjana R, Rio J, Auger C, et al. Brainstem lesions in clinically isolated syndromes. Neurology. 2010;75(21):1933-8.
212. Buuren Sv, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations inR. Journal of Statistical Software. 2011;45(3):1-67.
213. Dobson R, Dassan P, Roberts M, Giovannoni G, Nelson-Piercy C, Brex PA. UK consensus on pregnancy in multiple sclerosis: 'Association of British Neurologists' guidelines. Pract Neurol. 2019;19(2):106-14.

214. Vollmer B, Ontaneda D, Bandyopadhyay A, Cohn S, Nair K, Sillau S, et al. Discontinuation and comparative effectiveness of dimethyl fumarate and fingolimod in 2 centers. Neurol Clin Pract. 2018;8(4):292-301.
215. D'Amico E, Zanghi A, Callari G, Borriello G, Gallo A, Graziano G, et al. Comparable efficacy and safety of dimethyl fumarate and teriflunomide treatment in Relapsing-Remitting Multiple Sclerosis: an Italian real-word multicenter experience. Ther Adv Neurol Disord. 2018;11:1756286418796404.
216. Giuliani M, Logoteta, A., Prosperini, L. et al. Baseline characteristics associated with NEDA-3 status in fingolimod-treated patients with relapsing-remitting multiple sclerosis. Mult Scler Demyelinating Disord. 2017;2(10).
217. Jaklin AK, Benjaminsen E, Alstadhaug KB. Effectiveness of Natalizumab in Achieving No Evidence of Disease Activity (NEDA-3)-Data From a Local Norwegian Cohort. Front Neurol. 2021;12:765837.
218. Fernandez-Diaz E, Perez-Vicente JA, Villaverde-Gonzalez R, Berenguer-Ruiz L, Candeliere Merlicco A, Martinez-Navarro ML, et al. Real-world experience of ocrelizumab in multiple sclerosis in a Spanish population. Ann Clin Transl Neurol. 2021;8(2):385-94.
219. Cellerino M, Boffa G, Lapucci C, Tazza F, Sbragia E, Mancuso E, et al. Predictors of Ocrelizumab Effectiveness in Patients with Multiple Sclerosis. Neurotherapeutics. 2021;18(4):2579-88.
220. Cree BAC, Pradhan A, Pei J, Williams MJ, Opera I, investigators OIc. Efficacy and safety of ocrelizumab vs interferon beta-1a in participants of African descent with relapsing multiple sclerosis in the Phase III OPERA I and OPERA II studies. Mult Scler Relat Disord. 2021;52:103010.
221. Weinstock-Guttman B, Ramanathan M, Hashmi K, Abdelrahman N, Hojnacki D, Dwyer MG, et al. Increased tissue damage and lesion volumes in African Americans with multiple sclerosis. Neurology. 2010;74(7):538-44.
222. Perumal J, Fox RJ, Balabanov R, Balcer LJ, Galetta S, Makh S, et al. Outcomes of natalizumab treatment within 3 years of relapsing-remitting multiple sclerosis diagnosis: a prespecified 2-year interim analysis of STRIVE. BMC Neurol. 2019;19(1):116.
223. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort. JAMA Neurol. 2015;72(2):152-8.
224. Ozakbas S, Piri Cinar B, Baba C, Kosehasanogullari G, Sclerosis Research Group M. Self-injectable DMTs in relapsing MS: NEDA assessment at 10 years in a real-world cohort. Acta Neurol Scand. 2022;145(5):557-64.
225. Giovannoni G, Bermel R, Phillips T, Rudick R. A brief history of NEDA. Mult Scler Relat Disord. 2018;20:228-30.
226. Scolding N, Barnes D, Cader S, Chataway J, Chaudhuri A, Coles A, et al. Association of British Neurologists: revised (2015) guidelines for prescribing disease-modifying treatments in multiple sclerosis. Pract Neurol. 2015;15(4):273-9.
227. Scalfari A, Neuhaus A, Daumer M, Deluca GC, Muraro PA, Ebers GC. Early relapses, onset of progression, and late outcome in multiple sclerosis. JAMA Neurol. 2013;70(2):214-22.

228. Gasperini C, Prosperini L, Tintore M, Sormani MP, Filippi M, Rio J, et al. Unraveling treatment response in multiple sclerosis: A clinical and MRI challenge. Neurology. 2019;92(4):180-92.
229. Coupe P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging. 2008;27(4):425-41.
230. Commowick O, Wiest-Daessle N, Prima S. Automated diffeomorphic registration of anatomical structures with rigid parts: application to dynamic cervical MRI. Med Image Comput Comput Assist Interv. 2012;15(Pt 2):163-70.
231. Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage. 2002;17(1):479-89.
232. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310-20.
233. Valverde S, Cabezas M, Roura E, Gonzalez-Villa S, Pareto D, Vilanova JC, et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage. 2017;155:159-68.
234. Jha A, Oh C, Hesdorffer D, Diehl B, Devore S, Brodie MJ, et al. Sudden Unexpected Death in Epilepsy: A Personalized Prediction Tool. Neurology. 2021;96(21):e2627-e38.
235. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. 2008;2(4):1360-83.
236. O'Hagan A. On Outlier Rejection Phenomena in Bayes Inference. Journal of the Royal Statistical Society: Series B (Methodological). 1979;41(3):358-67.
237. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software. 2017;80(1):1-28.
238. Matthew D. Hoffman AG. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014;15:1593-623.
239. Vehtari A GA, Gabry J. Practical Bayesian model evaluation using leave-one-out crossvalidation and WAIC. Stat Comput Springer US. 2017;27:1413 - 32.
240. Oh J, Vidal-Jordana A, Montalban X. Multiple sclerosis: clinical aspects. Curr Opin Neurol. 2018;31(6):752-9.
241. Dobson R, Ramagopalan S, Davis A, Giovannoni G. Cerebrospinal fluid oligoclonal bands in multiple sclerosis and clinically isolated syndromes: a meta-analysis of prevalence, prognosis and effect of latitude. J Neurol Neurosurg Psychiatry. 2013;84(8):909-14.
242. Watanabe M, Nakamura Y, Sato S, Niino M, Fukaura H, Tanaka M, et al. HLA genotype-clinical phenotype correlations in multiple sclerosis and neuromyelitis optica spectrum disorders based on Japan MS/NMOSD Biobank data. Sci Rep. 2021;11(1):607.
243. Bross M, Hackett M, Bernitsas E. Approved and Emerging Disease Modifying Therapies on Neurodegeneration in Multiple Sclerosis. Int J Mol Sci. 2020;21(12).
244. Durelli L, Verdun E, Barbero P, Bergui M, Versino E, Ghezzi A, et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple

sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). Lancet. 2002;359(9316):1453-60.
245. Mikol DD, Barkhof F, Chang P, Coyle PK, Jeffery DR, Schwid SR, et al. Comparison of subcutaneous interferon beta-1a with glatiramer acetate in patients with relapsing multiple sclerosis (the REbif vs Glatiramer Acetate in Relapsing MS Disease [REGARD] study): a multicentre, randomised, parallel, open-label trial. Lancet Neurol. 2008;7(10):903-14.
246. Roos I, Leray E, Frascoli F, Casey R, Brown JWL, Horakova D, et al. Delay from treatment start to full effect of immunotherapies for multiple sclerosis. Brain. 2020;143(9):2742-56.
247. Chisari CG, Comi G, Filippi M, Paolicelli D, Iaffaldano P, Zaffaroni M, et al. PML risk is the main factor driving the choice of discontinuing natalizumab in a large multiple sclerosis population: results from an Italian multicenter retrospective study. J Neurol. 2022;269(2):933-44.
248. Wingerchuk DM, Weinshenker BG. Disease modifying therapies for relapsing multiple sclerosis. BMJ. 2016;354:i3518.
249. Coyle PK, Shang S, Xiao Z, Dong Q, Castrillo-Viguera C. Matching-adjusted comparisons demonstrate better clinical outcomes with SC peginterferon beta-1a every two weeks than with SC interferon beta-1a three times per week. Mult Scler Relat Disord. 2018;22:134-8.
250. Thomas Scott OM, Carmen Castrillo-Viguera, Adrian Harrington, Maria L. Naylor,. Peginterferon beta-1a every 2 weeks demonstrated better clinical outcomes than glatiramer acetate once-daily in patients with RRMS: Propensity score matching of phase 3 data from ADVANCE and CONFIRM. Revue Neurologique. 2019;175:S96.
251. Newsome SD, Mokliatchouk O, Castrillo-Viguera C, Naylor ML. Matching-adjusted comparisons demonstrate better clinical outcomes in patients with relapsing multiple sclerosis treated with peginterferon beta-1a than with teriflunomide. Mult Scler Relat Disord. 2020;40:101954.
252. Ontaneda D, Nicholas J, Carraro M, Zhou J, Hou Q, Babb J, et al. Comparative effectiveness of dimethyl fumarate versus fingolimod and teriflunomide among MS patients switching from first-generation platform therapies in the US. Mult Scler Relat Disord. 2019;27:101-11.
253. Boster A, Nicholas J, Wu N, Yeh WS, Fay M, Edwards M, et al. Comparative Effectiveness Research of Disease-Modifying Therapies for the Management of Multiple Sclerosis: Analysis of a Large Health Insurance Claims Database. Neurol Ther. 2017;6(1):91-102.
254. Reder AT, Arndt N, Roman C, Geremakis C, Mendoza JP, Su R, et al. Real-world propensity score comparison of treatment effectiveness of peginterferon beta-1a vs. subcutaneous interferon beta-1a, glatiramer acetate, and teriflunomide in patients with relapsing-remitting multiple sclerosis. Mult Scler Relat Disord. 2021;51:102935.
255. Cohen JA, Trojano M, Mowry EM, Uitdehaag BM, Reingold SC, Marrie RA. Leveraging real-world data to investigate multiple sclerosis disease behavior, prognosis, and treatment. Mult Scler. 2020;26(1):23-37.

256. Bose G, Rush C, Atkins HL, Freedman MS. A real-world single-centre analysis of alemtuzumab and cladribine for multiple sclerosis. Mult Scler Relat Disord. 2021;52:102945.
257. Zhu C, Zhou Z, Roos I, Merlo D, Kalincik T, Ozakbas S, et al. Comparing switch to ocrelizumab, cladribine or natalizumab after fingolimod treatment cessation in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2022.
258. Salter A, Lancia S, Cutter G, Marrie RA, Mendoza JP, Lewin JB, et al. A propensity-matched comparison of long-term disability worsening in patients with multiple sclerosis treated with dimethyl fumarate or fingolimod. Ther Adv Neurol Disord. 2021;14:17562864211021177.
259. Turner B, Cree BAC, Kappos L, Montalban X, Papeix C, Wolinsky JS, et al. Ocrelizumab efficacy in subgroups of patients with relapsing multiple sclerosis. J Neurol. 2019;266(5):1182-93.
260. Vollmer BL, Nair K, Sillau S, Corboy JR, Vollmer T, Alvarez E. Rituximab versus natalizumab, fingolimod, and dimethyl fumarate in multiple sclerosis treatment. Ann Clin Transl Neurol. 2020;7(9):1466-76.
261. Boremalm M, Juto A, Axelsson M, Novakova L, Frisell T, Svenningsson A, et al. Natalizumab, rituximab and fingolimod as escalation therapy in multiple sclerosis. Eur J Neurol. 2019;26(8):1060-7.
262. Granqvist M, Boremalm M, Poorghobad A, Svenningsson A, Salzer J, Frisell T, et al. Comparative Effectiveness of Rituximab and Other Initial Treatment Choices for Multiple Sclerosis. JAMA Neurol. 2018;75(3):320-7.
263. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult Scler. 2017;23(5):711-20.
264. Brownlee WJ, Altmann DR, Alves Da Mota P, Swanton JK, Miszkiel KA, Wheeler-Kingshott CG, et al. Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. Mult Scler. 2017;23(5):665-74.
265. Chen Y, Haacke EM, Bernitsas E. Imaging of the Spinal Cord in Multiple Sclerosis: Past, Present, Future. Brain Sci. 2020;10(11).
266. Lublin FD, Häring DA, Ganjgahi H, Ocampo A, Hatami F, Čuklina J, et al. How patients with multiple sclerosis acquire disability. Brain. 2022;145(9):3147-61.
267. Harding K, Williams O, Willis M, Hrastelj J, Rimmer A, Joseph F, et al. Clinical Outcomes of Escalation vs Early Intensive Disease-Modifying Therapy in Patients With Multiple Sclerosis. JAMA Neurol. 2019;76(5):536-41.
268. Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, et al. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. Sci Rep. 2020;10(1):9219.
269. Jain N, Bhatti MT. Fingolimod-associated macular edema: incidence, detection, and management. Neurology. 2012;78(9):672-80.
270. Hoffmann S, Cepok S, Grummel V, Lehmann-Horn K, Hackermuller J, Stadler PF, et al. HLA-DRB1*0401 and HLA-DRB1*0408 are strongly associated with the development of antibodies against interferon-beta therapy in multiple sclerosis. Am J Hum Genet. 2008;83(2):219-27.

271. Tur C, Ramagopalan S, Altmann DR, Bodini B, Cercignani M, Khaleeli Z, et al. HLA-DRB1*15 influences the development of brain tissue damage in early PPMS. Neurology. 2014;83(19):1712-8.
272. Brownlee WJ, Tur C, Manole A, Eshaghi A, Prados F, Miszkiel KA, et al. HLA-DRB1*1501 influences long-term disability progression and tissue damage on MRI in relapse-onset multiple sclerosis. Mult Scler. 2023;29(3):333-42.
273. Benedict RH, Amato MP, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. BMC Neurol. 2012;12:55.
274. Van Schependom J, D'Hooghe M B, Cleynhens K, D'Hooge M, Haelewyck MC, De Keyser J, et al. The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis. Eur J Neurol. 2014;21(9):1219-25, e71-2.
275. Kavaliunas A, Tinghog P, Friberg E, Olsson T, Alexanderson K, Hillert J, et al. Cognitive function predicts work disability among multiple sclerosis patients. Mult Scler J Exp Transl Clin. 2019;5(1):2055217318822134.
276. Skorve E, Lundervold AJ, Torkildsen O, Myhr KM. A two-year longitudinal follow-up of cognitive performance assessed by BICAMS in newly diagnosed patients with MS. Mult Scler Relat Disord. 2020;46:102577.
277. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med. 2021;27(5):815-9.

Chapter 8. Appendix

8.1 STROBE checklist for chapter three

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	93
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	N/A
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	93-95
Objectives	3	State specific objectives, including any prespecified hypotheses	95
Methods			
Study design	4	Present key elements of study design early in the paper	96
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	96-100
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	96-97
		(b) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	99-100
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	99-100
Bias	9	Describe any efforts to address potential sources of bias	96
Study size	10	Explain how the study size was arrived at	96-97
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	97-100
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	97-98
		(b) Describe any methods used to examine subgroups and interactions	N/A
		(c) Explain how missing data were addressed	98
		(d) If applicable, explain how loss to follow-up was addressed	103, 106, 110, 113, 117
		(e) Describe any sensitivity analyses	N/A
Results			
Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	101-117

		(b) Give reasons for non-participation at each stage	118-128
		(c) Consider use of a flow diagram	Tables 36, p 119
Descriptive data	14	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	101-117
		(b) Indicate number of participants with missing data for each variable of interest	101-117
		(c) Summarise follow-up time (eg, average and total amount)	101-117
Outcome data	15	Report numbers of outcome events or summary measures over time	101-128
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	101-128
		(b) Report category boundaries when continuous variables were categorized	N/A
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	129-132
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	129-132
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	129-132
Generalisability	21	Discuss the generalisability (external validity) of the study results	129-132
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	96

8.2 STROBE checklist for chapter four

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	133
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	N/A
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	133
Objectives	3	State specific objectives, including any prespecified hypotheses	133
Methods			
Study design	4	Present key elements of study design early in the paper	134-136
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	134
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	134
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	134-135
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	134-136
Bias	9	Describe any efforts to address potential sources of bias	135-136
Study size	10	Explain how the study size was arrived at	134-135
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	134-135
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	134-140
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	140
		(*d*) If applicable, explain how loss to follow-up was addressed	140
		(*e*) Describe any sensitivity analyses	N/A
Results			

Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	142-144
		(b) Give reasons for non-participation at each stage	142-144
		(c) Consider use of a flow diagram	Table 49
Descriptive data	14	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	142-144
		(b) Indicate number of participants with missing data for each variable of interest	142-144
		(c) Summarise follow-up time (eg, average and total amount)	142-144
Outcome data	15	Report numbers of outcome events or summary measures over time	142-144
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	142-149
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and **sensitivity analyses**	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	150
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	150-152
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	150-152
Generalisability	21	Discuss the generalisability (external validity) of the study results	150-152
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A

8.3 STROBE checklist for chapter five

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	153
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	N/A
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	153
Objectives	3	State specific objectives, including any prespecified hypotheses	153-154
Methods			
Study design	4	Present key elements of study design early in the paper	154
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	154
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	154
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	168
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	154
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	154
Bias	9	Describe any efforts to address potential sources of bias	154
Study size	10	Explain how the study size was arrived at	154
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	154
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	155
		(*b*) Describe any methods used to examine subgroups and interactions	155
		(*c*) Explain how missing data were addressed	156
		(*d*) If applicable, explain how loss to follow-up was addressed	N/A
		(*e*) Describe any sensitivity analyses	155
Results			
Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	154-156
		(b) Give reasons for non-participation at each stage	156
		(c) Consider use of a flow diagram –	156

Descriptive data	14	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	154-155
		(b) Indicate number of participants with missing data for each variable of interest	156
		(c) Summarise follow-up time (eg, average and total amount)	156
Outcome data	15	Report numbers of outcome events or summary measures over time	156-163
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	156-166
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	167
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	167-172
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	167-172
Generalisability	21	Discuss the generalisability (external validity) of the study results	167-172
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A