Multiple Sclerosis MSBase registry: Using real-world data to define MS outcomes and optimise treatment strategies

Timothy Denis Spelman

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

February 2016

Department of Medicine

The University of Melbourne
Abstract

Multiple sclerosis (MS) is a progressive, chronic and inflammatory demyelinating disease of the central nervous system that represents a significant social and economic burden to sufferers and the community. Whilst primary evidence from clinical trials supports the short-term efficacy and safety of MS-specific treatments, the typical 12 or 24-month trial period captures only a small proportion of a patient's overall MS experience which can span decades. A considerable challenge to MS epidemiology and treatment research involves developing and adapting statistical techniques for separating out genuine signals from the considerable noise which often characterises data collected from patients with a complex, long-term disease such as MS. Analysis of appropriately powered, longitudinal real world databases is central to providing novel insights into both the biological mechanisms of disease and better targeting and management of treatment for a debilitating and chronic disease for which there is presently no cure.

This thesis details a range of novel applications and adaptations of a variety of statistical techniques for analysing real world MS outcomes using the international MSBase registry of multiple sclerosis and comparing a range of competing treatment strategies for which there currently exists limited precedent data to guide decision making in clinical practice. A series of analytical themes ranging from global epidemiology through to the derivation of personalised risk prediction tools and conduct of propensity-score matched head-to-head treatment and management comparisons are presented and these illustrate the flexibility and utility of a large, global registry data source.

Amongst the various analyses presented herein, this report presents a novel use of trigonometric regression modelling to demonstrate for the first time that there is a latitude-dependent relationship between seasonal ultraviolet radiation (UVR) trough and relapse frequency peak independent of location-specific UVR levels, with latitudes further from the equator associated with...
shorter gaps. This modelling exercise further confirms prior meta-analyses showing a strong seasonal relapse onset probability variation in the northern hemisphere, and extends this observation, again for the first time, to the southern hemisphere. An analysis of the, to date, largest ever studied seen-from-onset MS cohort is also presented, culminating in the derivation of a series of empirically grounded prognostic nomograms for personalised risk assessment of conversion to clinically definite MS in clinical practice. Finally a variety of propensity score matching methods are tested and applied to a series of head-to-head treatment comparisons and decision points, including initiation, switch and discontinuation triggers, to return relatively unbiased estimation of comparative treatment efficacy in clinically relevant scenarios using non-randomly assigned observational data.
Declaration

The following declaration page, signed by the candidate:

This is to certify that:

i. the thesis comprises only my original work towards the PhD,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.
Preface

The structure and operating conditions of the MSBase registry provides co-authorship on the basis of the academic effort involved in collecting, collating and uploading patient data into the registry in addition to engagement in the review of draft manuscript once analysis is complete and the associated paper drafted. Given the MSBase rules for authorship are largely based on the quantity of data contributed to a particular analysis, the list of eligible authors is typically long, changes with every analysis and sourced from multiple clinics and countries. As such, a summary of author contributions are provided separately for each analysis as appropriate.

For all analyses described herein, Tim Spelman was solely responsible for data formatting, cleaning and preparation once extracted from the registry, programming, conducting the analysis, compiling and writing the reports and fully managing the revision process. Design, scoping and feasibility for each analysis presented were undertaken in collaboration with Helmut Butzkueven, the primary thesis supervisor and senior co-author on each of the papers and reports presented herein. To reflect this, signed co-authorship forms are provided for Professor Butzkueven’s contribution in each of the published or in-print papers or reports.
Acknowledgements

The author acknowledges the support, guidance and expertise of supervisors Helmut Butzkueven and Danny Liew and further expresses gratitude to Stephen Rogerson for his advice. Further analysis-specific acknowledgements are provided within each of the presented papers and reports herein.
CONTENTS (page number)

Title page........................................................................................................................................................................1
Abstract.............................................................................................................................................................................2
Declaration........................................................................................................................................................................4
Preface...............................................................................................................................................................................5
Acknowledgements.............................................................................................................................................................6
Contents.............................................................................................................................................................................7
List of tables....................................................................................................................................................................12
List of figures....................................................................................................................................................................14

A. INTRODUCTION.............................................................................................................................................................16
  A.1 Thesis questions.........................................................................................................................................................22
  A.2 Thesis structure............................................................................................................................................................23
  A.3 References..................................................................................................................................................................24

B. LITERATURE REVIEW......................................................................................................................................................28
  B.1 Hypothesis 1: The timing of relapse onset is seasonal and this relationship is latitude dependent..............28
    B.1.1 Search Strategy...................................................................................................................................................28
    B.1.2 Results..................................................................................................................................................................28
      B.1.2.1 Seasonality and relapse.................................................................................................................................29
      B.1.2.2 Seasonality and non-relapse MS activity........................................................................................................35
      B.1.2.3 Seasonality, birth month and MS onset risk.................................................................................................37
      B.1.2.4 Latitude and MS activity................................................................................................................................39
      B.1.2.5 Seasonality and ultraviolet radiation...............................................................................................................41
    B.1.3 Conclusion............................................................................................................................................................42
    B.1.4 References............................................................................................................................................................42
B.2 Hypothesis 2: Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis........46

B.2.1 Search strategy.................................................................................................................................................46

B.2.2 Results..................................................................................................................................................................46

B.2.2.1 MRI..................................................................................................................................................................47

B.2.2.2 Disease-modifying drugs...............................................................................................................................51

B.2.2.3 Cerebrospinal fluid examination.....................................................................................................................53

B.2.2.4 Biomarkers......................................................................................................................................................55

B.2.2.5 Developmental and environmental correlates...............................................................................................57

B.2.2.6 Methods for isolating independent predictors of conversion........................................................................58

B.2.2.7 Methods for identifying risk sub-groups.......................................................................................................61

B.2.3 Conclusion..........................................................................................................................................................62

B.2.4 References........................................................................................................................................................63

B.3 Hypothesis 3: Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict early discontinuation........................................................................68

B.3.1 Search strategy..................................................................................................................................................68

B.3.2 Results..................................................................................................................................................................68

B.3.2.1 Frequentist predictor analyses.........................................................................................................................69

B.3.2.2 Non-frequentist methods..................................................................................................................................73

B.3.3 Conclusion..........................................................................................................................................................74

B.3.4 References........................................................................................................................................................75

B.4 Hypothesis 4: Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment scenarios and products.........................................................................78

B.4.1 Search strategy..................................................................................................................................................78

B.4.2 Results..................................................................................................................................................................78

B.4.2.1 Propensity score matching methodology & performance..................................................................................79

B.4.2.2 Propensity score adjustment in multiple sclerosis.............................................................................................83

B.4.2.3 Reporting and evaluation..................................................................................................................................91

B.4.3 Conclusion..........................................................................................................................................................93
B.4.4 References......................................................................................................................................................93

C. METHODS...............................................................................................................................................................98

C.1 The timing of relapse onset is seasonal and this relationship is latitude-dependent – methods..........98

C.1.1 A method of trigonometric modelling of seasonal variation - the example of multiple sclerosis relapse
data paper).................................................................................................................................................................98

C.1.1.1 Protocol code (published file)..........................................................................................................................127

C.2 Quantifying risk of early relapse in high risk patients with first demyelinating events – methods.......137

C.2.1 Background..........................................................................................................................................................137

C.2.1 Nomogram construction.........................................................................................................................................137

C.2.2 Nomogram interpretation..........................................................................................................................................140

C.2.3 Nomogram validation..............................................................................................................................................141

C.2.3.1 Concordance index................................................................................................................................................141

C.2.3.2 Calibration..........................................................................................................................................................142

C.2.4 References............................................................................................................................................................144

C.3 Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a
range of treatment settings and products – methods......................................................................................................145

C.3.1 Background..........................................................................................................................................................145

C.3.1.1 References........................................................................................................................................................145

C.3.2 Binomial propensity score matching – 1:1 matching..............................................................................................145

C.3.3 Binomial propensity score matching – many-to-one matching.............................................................................148

C.3.3.1 References........................................................................................................................................................150

C.3.4 Multinomial propensity score matching across three treatment groups.........................................................151

C.3.4.1 References........................................................................................................................................................153

C.3.5 Post-matching assessment of balance..................................................................................................................154

C.3.5.1 Paired tests........................................................................................................................................................154

C.3.5.2 Standardised differences.....................................................................................................................................154

C.3.5.2.1 References.......................................................................................................................................................155
D. ANALYSIS - The timing of relapse onset is seasonal and this relationship is latitude-dependent

D.1 Seasonal variation of relapse rate in multiple sclerosis is latitude-dependent (paper).........161

E. ANALYSIS - Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis

E.1 Quantifying risk of early relapse in patients with first demyelinating events........................................192

F. ANALYSIS - Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict early discontinuation

F.1 Gender, country, baseline EDSS and on-treatment relapse rate predict first DMT treatment discontinuation.................................................................224

G. ANALYSIS - Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment settings and products

G.1 Comparative efficacy of switching to natalizumab in active relapsing multiple sclerosis.......................249

G.2 Comparative efficacy of first-line natalizumab versus IFNβ or glatiramer acetate in relapsing MS.................................................................287

G.3 Risk of early relapse following switch from injectables to oral agents for multiple sclerosis.................316
G.4 Comparison of efficacy and persistence of first line fingolimod vs interferon-beta/glatiramer in the presence of prior disease activity..........................................................339
G.5 Comparison of Switching to Natalizumab versus Remaining on Interferon-Beta or Glatiramer Acetate after On-Treatment MS Relapse Using Propensity-Matched Registry Data (paper).................................366
G.6 Comparative analysis of MS disability regression independent of relapse recovery in natalizumab-treated patients using multinomial propensity score matching.................................................................388

H. FINAL DISCUSSION.................................................................................................................404

H.1 Summary of findings...........................................................................................................404
H.2 MS registry data – advantages, opportunities and future directions.................................408
H.3 Limitations..........................................................................................................................413
H.4 Conclusion..........................................................................................................................414

I. APPENDIX A............................................................................................................................415
I.1 Hypothesis 1 literature review – PRISMA flowchart............................................................415
I.2 Hypothesis 2 literature review – PRISMA flowchart............................................................416
I.3 Hypothesis 3 literature review – PRISMA flowchart............................................................417
I.4 Hypothesis 4 literature review – PRISMA flowchart............................................................418

J. APPENDIX B............................................................................................................................419

J.1 List of related publications and conference proceedings.......................................................419
List of Tables (page number)

Table D.1.1: Patient characteristics.................................................................173
Table E.1.1: Demography, disease, treatment and examination characteristics.............................................201
Table E.1.2: Predictors of first post-CIS relapse.......................................................................................202
Table E.1.3: Predictors of first post-CIS relapse – by DMD exposure status..............................................208
Table E.1.s1: Predictors of first post-CIS relapse – alternate model substituting T2 lesions on spinal MRI for oligoclonal bands on CSF examination...............................................................219
Table E.1.s2: Sensitivity analysis: Predictors of first post-CIS relapse – limiting primary model to subset of patients with baseline MRI within 3 months of CIS..........................................................220
Table F.1.1: Demography, disease and treatment characteristics......................................................................233
Table F.1.2: Predictors of treatment discontinuation – first DMD.................................................................236
Table F.1.3: Predictors of first DMD treatment discontinuation using different censor points...................239
Table F.1.4: Predictors of treatment discontinuation – any DMY and post-first DMT........................................241
Table G.1.1: Pre-matching comparison of baseline characteristics by switch group..............................262
Table G.1.2: Propensity-matching comparison of baseline characteristics by switch group...................263
Table G.1.3: Annualised relapse rates by treatment groups and post-baseline year.......................................265
Supplementary table G.1.1: Baseline characteristics of unmatched patients by switch group.............286
Table G.2.1: First-line natalizumab – baseline characteristics of unmatched patients...............................303
Table G.2.2: First-line natalizumab – baseline characteristics of propensity score-matched patients..........................................................303
Table G.2.3: Summary of propensity-matched treatment group outcomes by prior disease activity..........................305
Table G.3.1: Switch vs stay – baseline characteristics of unmatched patients..............................................327
Table G.3.2: Switch vs stay – baseline characteristics of propensity score-matched patients..................328
Table G.4.1: Comparison of baseline characteristics – unmatched..........................................................350
Table G.4.2: Comparison of baseline characteristics in the 2:1 propensity matched patients.................350
**Table G.5.1:** IFNβ/GA stay vs switch - comparison of baseline characteristics by unmatched treatment groups..................................................................................................................................................375

**Table G.5.2:** IFNβ/GA stay vs switch - comparison of baseline characteristics by 2:1 propensity matched treatment groups..................................................................................................................................................376

**Table G.6.1:** Multinomial PS matching - comparison of baseline characteristics across the unmatched treatment switch groups..................................................................................................................................................394

**Table G.6.2:** Multinomial PS matching - comparison of baseline characteristics across the propensity matched treatment switch groups..................................................................................................................................................395

**Table G.6.3:** Comparison of ARR across the three matched treatment switch groups..................395
List of Figures (page number)

**Figure C.1.1.1:** Radar plots of observed global relapse frequency by month in the 1a) northern hemisphere, 1b) combined northern and southern hemispheres.................................................................113

**Figure C.1.1.2:** Plot comparing observed monthly relapses by hemisphere with predicted relapses using the base-case geometric model describing a single annual cycle of one peak and one trough separated by six months..................................................................................................................114

**Figure C.1.1.3:** Plots comparing observed median monthly UVR with base model predicted UVR for Montreal, Canada; Melbourne, Australia; Bari, Italy & Buenos Aires, Argentina.............................................115

**Figure C.1.1.4:** Weighted line of best fit between absolute latitude and UV-trough-to-relapse-peak lag ..................................................................................................................................................117

**Figure C.4.2.1:** Example of a cumulative AUC measurement from a sample 24-month EDSS/time plot.................................................................................................................................................158

**Figure D.1.1:** Sine-wave regression of number relapse onsets by month in the northern and southern hemispheres.................................................................................................................................175

**Figure D.1.2:** Spider-web plot of distribution of monthly relapses combined across both hemispheres..................................................................................................................................................175

**Figure D.1.3:** Weighted line of best fit between latitude and UV trough-to-relapse peak...............................................................177

**Figure E.1.1:** Kaplan-Meier survival curve – time to second attack by DMD exposure status........203

**Figure E.1.2:** Nomogram for 12-month conversion.................................................................................................................................205

**Figure E.1.s1:** Worked example of how to interpret nomogram.................................................................................................................206

**Figure E.1.3:** Calibration curve for 12-month conversion..........................................................................................................................207

**Figure E.1.s2:** Nomogram for 6-month conversion...............................................................................................................................221

**Figure E.1.s3:** Nomogram for 2-year conversion.................................................................................................................................221

**Figure E.1.s4:** Nomogram for 3-year conversion.................................................................................................................................222

**Figure E.1.s5:** Nomogram for 4-year conversion.................................................................................................................................222
Timothy Denis Spelman (58172)

Figure E.1.s6: Nomogram for 5-year conversion.................................................................223

Figure F.1.1: Time to first DMD discontinuation by product identity........................................237

Figure G.1.1: Natalizumab switch analysis – study profile..........................................................256

Figure G.1.2: Example of a cumulative AUC measurement from a sample 24-month EDSS/time plot................................................................................................................................................260

Figure G.1.3: Time to A) first relapse, B) treatment discontinuation, or C) three-month confirmed disability progression after treatment switch..........................................................................................................................266

Figure G.1.4: Time to first relapse after treatment switch by prior treatment subgroup, (A) IFNβ, (B) GA, or (C) IFNβ and GA.........................................................................................................................................................270

Figure G.1.5: Time to treatment discontinuation after treatment switch by prior treatment subgroup, (A) IFNβ, (B) GA, or (C) IFNβ and GA.........................................................................................................................................................271

Supplemental Figure G.1.1: Distribution of propensity scores by treatment arm prior to (A) and after (B) propensity matching.........................................................................................................................................................285

Figure G.2.1: First line natalizumab vs BRACE – study profile..................................................299

Figure G.2.2: Time to A) first relapse and (B) treatment discontinuation on first-line treatment.....304

Figure G.3.1: Time to first relapse within 6 months post-baseline.................................................329

Figure G.3.2: Sensitivity analysis: Time to any first post-baseline relapse....................................331

Figure G.4.1: Fingolimod vs BRACE; Time to first relapse..........................................................352

Figure G.4.2: Fingolimod vs BRACE; Time to confirmed disability progression..........................353

Figure G.5.1: IFNβ/GA stay vs switch - time to first relapse (simultaneous censoring).................377

Figure G.5.2: IFNβ/GA stay vs switch - time to treatment discontinuation.....................................378

Figure G.5.3: IFNβ/GA stay vs switch - time to three-month confirmed disability progression........379

Figure G.6.1: Multinomial PS matching - time to first post-switch relapse by switch treatment group.........................................................................................................................................................396

Figure G.6.2: Multinomial PS matching - time to first post-switch disability progression event by switch treatment group.........................................................................................................................................................397
A. Introduction

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system. Around 90-95% of people with MS first experience focal episodes of neurological impairment followed by a recovery phase. The first of these events is conventionally referred as either “first demyelinating event (FDE)” or “clinically isolated syndrome (CIS)”. Subsequent clinical episodes are termed relapses and the first clinically confirmed relapse changes the patient classification to relapsing-remitting MS, provided a series of other inclusion/exclusion criteria are met, which have varied over time as different diagnostic criteria evolve. The relapsing-remitting phase of MS is often followed, a decade or more later, by a secondary progressive phase characterised by a gradual accumulation of disability (e.g., typically a progressive gait disturbance with other progressive neurological impairments). In Australia, MS prevalence is estimated at between 89.3 cases per 100,000 persons to 95.6 per 100,000, equating to approximately 23,000 patients diagnosed nationally. This compares with an estimate of two million patients internationally. The average age of first symptoms is characteristically between 20 and 40 years and a disproportionate number of women are affected (approximately 75%). Total cost of MS in the Australian setting was estimated in 2010 to be $48,945 per patient per annum, with indirect costs (i.e., those not related to treatment or care) associated with the loss of earning potential capturing the largest share of these costs. This, coupled with the relatively young age at diagnosis means that MS represents a considerable economic burden over a wide time horizon at both the level of the individual patient and broader community and health care system. This is in addition to the marked social burden and erosion in quality of life for sufferers.

The cause of MS is unknown. Genetic and environmental modulators have been postulated although no single underlying biological mechanism has thus far been established. It is probable that causation emerges, perhaps stochastically, from a complex and multifactorial interaction of genetic
and environmental factors. A considerable challenge to MS epidemiology involves isolating independent predictors of MS onset, relapse and progression. As such developing and adapting statistical techniques for separating out genuine signals from the considerable noise which often characterises data collected from patients with a complex disease such as MS is central to providing novel insights into the biological mechanisms of disease and, by extension, better targeting of effective and better tolerated treatments for a debilitating and chronic disease for which currently there is no cure. An example, illustrated in this thesis is the confirmation of a strong seasonal variation of relapse onset probability. One strong biological hypothesis is a UV-related immunomodulatory effect, and resultant UV-related therapeutic strategies (UV therapy or vitamin D supplementation) are currently being examined in large randomised trials.

Central to robust analysis is the availability of high quality data. Large, longitudinal MS registries therefore present a valuable tool for studying such complex phenomena in a real-world setting. Whilst there is a solid randomised clinical trial evidence base in MS supporting the efficacy of disease-modifying drugs (DMD) for the prevention or delay of relapse or disability progression, at least relative to placebo, these studies are typically limited to 12 or 24 months duration. Whilst a limited number of open label extension studies have permitted longer views of, primarily, the relative efficacy of early vs delayed treatment, none of these have been able to follow patients over the much longer disease durations experienced by the majority of patients. Indeed, the 2010 Hurwitz expert panel review of MS registries recommended that data from appropriately analysed registry-based (MS) studies could be combined with the results coming from clinical trials to both optimise treatment and, in particular, improve long term patient outcomes. This is particularly timely and relevant given the recent acceleration in development, trials and approval of novel agents for MS treatment.
Registries for MS have existed for at least the past 50-60 years. The systematic collection and databasing of information on MS patients date back to the late 1950s, when an early database predecessor for what ultimately evolved in the early nineties as the Danish MS scientific registry was established in 1956. Database-style registries were subsequently established in Canada including the much-studied London, Ontario MS database, established in 1972 and, later, the British Columbia Multiple Sclerosis database, first set-up in 1980 using the MS Computer Stored Ambulatory Record (COSTAR) database. This was later followed in the mid-1990s by a number of US based registries including the clinician-driven North American Research Committee on Multiple Sclerosis (NARCOMS) registry, the New York State Multiple Sclerosis Consortium (NYSMSC) database in 1996, and the service/usage oriented Department of Veterans Affairs MS surveillance registry. Additional European-based databases soon followed throughout the 1990s and then the 2000s. These included the Oslo Multiple Sclerosis Registry, enrolling prospective MS patients since 1990, the French-based, clinician-managed European Database for Multiple Sclerosis (EDMUS), the closely related Swedish Interactive Database for Multiple Sclerosis (IDMS), a forerunner of the later web-based national Swedish MS registry, and the large Italian Multiple Sclerosis Database Network (MSDN), the first of its kind at the time in Italy. More recently newer registries are currently taking shape in Germany, Poland, Spain and the UK.

As database functionality and internet connectivity have evolved, the collation, storage and sharing of MS data is no longer limited to what were previously typically site or country only registries. These have subsequently given way to truly international registries such as the MSBase registry which provides new opportunities to study MS across borders and settings. This includes the study of epidemiological phenomena such as seasonal and latitudinal gradients and the ability to accumulate large quantities of treatment exposure records to study long term DMD effects with appropriate power. By comparison, a limitation of collating or synthesising data from separate, typically single-setting registry databases is the significant between-database heterogeneity.
secondary to systematic differences in patient characteristics, definitions used and reporting protocols. Whilst pooling data from databases can be an attractive option for studying MS outcomes that are known or suspected of requiring large samples and power, the often significant lack of exchangeability between MS registries can limit such analyses to naïve, indirect comparisons only due to the insufficient homogeneity required for formal pooling or other such meta-analytical or meta-regression methods. Indeed the Hurwitz review of MS registries further recognised this as considerable limitation to progressing appropriately powered MS real world outcomes research, recommending alignment of and consistency within data definitions used and responses collected across MS registries. Whilst the minimum datasets of the French, Italian, Swedish and indeed the international MSBase registry are closely aligned, systematic differences in the frequency of recording EDSS in clinical practice within US collections makes comparison difficult.  

A traditional function of disease registries has been to facilitate epidemiological studies, owing to, in part, the wealth of data collected over typically long periods of time across often broad geographical areas. As such they can be particularly adept at studying long term disease trajectories or disease outcomes requiring years, and in the case of MS decades, to declare themselves including disability progression and environmental correlates of MS risk. Population based MS registries such as the Danish MS registry have been an important tool for flagging or, equally, debunking candidate environmental risk factors for MS including season of birth, latitude, infection, head trauma, and even theorised occupational exposure to solvents. Long term follow up in epidemiological studies in MS is particularly important given the potentially large time intervals that may elapse between a suspected environmental exposure and subsequent MS event such as onset, relapse or disability progression. Similarly, broad geographical coverage is key to studying candidate environmental correlates of MS onset or activity risk that describe gradients in time and/or space such as latitude and seasonally-moderated influences of risk. Whilst strictly epidemiological studies of observational cohorts are, by definition, exploratory and not formal tests of an underlying
mechanism linking a particular environmental insult or exposure to later MS activity, registry-based exploration can be used to inform subsequent trials designed for examination of a causal link.

Beyond facilitating natural history studies and the analysis of long term outcomes, a registry approach for collating and tracking MS patient outcomes over time is a potentially important tool for clinicians and researchers in the study and isolation of independent predictors of poor outcome or, conversely, good response to treatment, optimal treatment sequencing, or favourable long-term safety profiles, all of which influence a clinician’s and patient’s decision to treat, or wait, and informs switch decisions. For example, analyses of long-term outcome data permits the identification of subsets of patients at higher risk for experiencing early relapse or progression event based on, for example, demographic, disease and examination characteristics at the time of first presentation with clinically isolated syndrome (CIS/FDE).  

Whilst MS data sourced from such registries provide opportunities to study long term outcomes in real world patient populations, their nature is non-randomised observational data, subject to the various limitations associated with analysing and extrapolating from observational cohorts including informed treatment selection, unmeasured confounding, differential follow-up and ascertainment and missing or incompletely observed data. The evidence base in MS, like many chronic diseases, is split between, on the one hand, a rich body of appropriately randomised clinical trials, particularly those concerning comparative treatment efficacy and, on the other hand a relative paucity of long-term real-world outcome studies. Whilst analysis of observational cohorts is no substitute for a controlled clinical trial, there are many scenarios in the study of chronic disease where clinical trials are simply not feasible, affordable or ethical, particularly when in the study of long term treatment outcomes or the long term performance of prognostic correlates of a poorer outcome. Whilst no replacement for formal registration studies, appropriately analysed data from observational databases can supplement that knowledge base and is increasingly demanded by regulators wishing
to be reassured that long-term treatment benefits are real rather than purely derived from complex predictive models and that serious long-term treatment risks, particularly incidence of malignancy and serious infection, are within expected ranges.

A systematic review of 136 articles between 1985 and 1998 and published in the year 2000 reporting the same head-to-head treatment comparisons in both an observational study and an equivalent randomised clinical trial reported that across the 19 treatment interventions analysed, in only 2 (10.5%) did the effect size estimated in the observational study fall outside the 95% confidence interval around the point estimate effect size observed in the equivalent clinical trial. This suggests that observational studies do not systematically over-estimate or under-estimate the true treatment effect as suggested in earlier reviews from the 1980s. Of course, how well an observational estimate approximates the true treatment effect is largely a function of the registry quality and the methodology employed to minimise selection bias and confounder imbalance, a focus of this thesis. Of note, the Hurwitz review observed that of the registries included in the review (which notably did not include either the large MSBase or Italian MSDN registries) were particularly limited in terms of treatment outcome data, concluding that none of the studied registries routinely collected sufficient treatment data to permit comparative analysis of the impact of competing treatments on the natural history of MS.

The primary data source for the analyses presented in this thesis is the international MSBase registry. The MSBase registry was established in 2004 as an on-line platform for the sharing, tracking and analysis of real-world international MS outcome data. The Registry functions as an international online database accumulator and collects disease related information from consenting patients attending contributing Multiple Sclerosis (MS) clinics. The registry is a collaborative research group that prospectively collects outcomes data from MS treatment centres using an internet-based, physician owned and operated system www.msbase.org. Each centre enters patient data in the
offline iMed© local electronic database during routine clinic visits and intermittently uploads codified datasets to the MSBase server. Physicians record clinical information such as date of MS onset, The Kurtzke Expanded Disability Status Score (EDSS) and functional system scores (a quantitated 8 dimension neurological examination and ambulation classifier), relapse characteristics, MRI and other investigations and diagnostic criteria used. Records are classified as complete and eligible for analyses if they meet a minimum required set of data. Quality of the EDSS assessment is assured by the requirement of online EDSS competency (Neurostatus) certification at each of the participating centres (www.neurostatus.net). Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing centre. The use of a unified data entry protocol, the requirement of a minimum dataset and minimum frequency of upload an assistance with neurostatus certification of contributing clinicians ensures a consistency and quality of data difficult to replicate via amalgamation of data from separate registry sources.\textsuperscript{26,39-41}

\textbf{A.1 Thesis questions}

This thesis uses data from the MSBase registry to study the following four hypotheses:

1) The timing of relapse onset is seasonal and this relationship is latitude-dependent

2) Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis

3) Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict rate of discontinuation
4) Propensity-score matching can return relatively unbiased estimates of comparative treatment efficacy across a range of treatment scenarios and disease-modifying drugs.

**A.2 Thesis structure**

This thesis is structured to present each of these four hypotheses and their associated analyses in order of an increasingly focused scale to progressively illustrate how the application of various statistical and mathematical techniques to the same data source can permit inferences around, at one end, global-scale, epidemiological MS gradients through to, at the other end of the scale pseudo-randomised, focused head-to-head treatment efficacy comparisons.

Section B presents the results of the literature review whilst section C details the statistical methods used for each analysis. Section D then describes a novel application of trigonometric modelling of seasonal and latitudinal correlates of relapse timing and probability. Section E explores prognostic correlates of time to conversion from clinically isolated syndrome (CIS) to clinically definite MS (CDMS) and presents a series of prognostic nomogram-based physician tools for personalised risk assessment in the clinical setting. Section F then investigates predictors of treatment persistence and, in particular, premature treatment discontinuation. The thesis concludes by presenting a series of propensity matched head-to-head efficacy comparisons (section G) across a range of disease-modifying drug (DMD) products through a range of common initiation and switching scenarios using a variety of matching techniques.

As the thesis covers four distinct, yet related, primary study questions, the literature review presented in the section B is split into four separate reviews – one for each primary hypothesis.
A.3 References


B. Literature review

Separate literature reviews were undertaken for all four of the primary thesis hypotheses. Summaries of each review are presented in order below.

B. 1 Hypothesis 1: The timing of relapse onset is seasonal and this relationship is latitude-dependent

B.1.1 Search Strategy

References for this review were identified using a search of the PubMed and Scopus electronic databases of peer-reviewed literature using a pre-specified review protocol. The search included the terms “multiple sclerosis” in combination with “seasonality” or “latitude” within either the article title or abstract published on or before 2013. The references of evaluated articles were then manually screened for additional publications that satisfied the search criteria. The Cochrane review library was then reviewed to identify relevant systematic reviews and meta-analyses not identified in the initial PubMed/Scopus search. The search included references in both English and languages other than English. Case reports were not reviewed. After removing duplicate records, the full reports of each reference satisfying the search criteria were reviewed for relevance and quality to arrive at the final reference. This search process is summarised in the PRISMA flow diagram (Appendix I.1).

B.1.2 Results

The initial screen of the electronic databases and manual review of references identified 53 references. A total of 16 records (31.4% of the initial screen) were rejected, 14 on the basis of insufficient relevance and/or quality and 2 as duplicates. Of the 37 papers satisfying the inclusion criteria five broad themes were identified – 1) seasonality and relapse, 2) seasonality and non-relapse MS activity, 3) seasonality, birth month and MS onset, 4) latitude and MS activity and 5) seasonality and ultraviolet radiation.
B.1.2.1 Seasonality and relapse

Of the total 35 references satisfying the search criteria, 15 (43%) directly studied seasonality in the timing of relapse onset in MS. Of these, the majority were single centre and/or single geographical location based studies only. Only two studies (5.7%) formally tested a hypothesis of an annual, seasonal cyclic trend in relapse prevalence across multiple countries. A 2013 retrospective analysis of thirty years of relapse data (1980-2010) sourced from seven sites across six countries (including one southern hemispheric location) from the MSBase registry was compared using weighted Peto odds ratios. Whilst a consistent summer-spring peak in relapse frequency was observed within all seven sites, the precise timing of this peak within this range varied by decade with earlier (1991-2000) relapse data tending to report a summer peak whilst later data (2001-2010) favoured a spring peak. Whilst the authors hypothesised that such a long-term shift in relapse seasonality may be consistent with an underlying mechanism favouring infectious correlates over UVR exposure, neither site-specific UVR nor other potential patient-level confounders of relapse timing were included in the modelling to better isolate the independent effect of season or test whether the observed shift in peak across decades is a genuine signal or an artefact of reporting or uncontrolled confounding including broader availability to a wider range of therapies.1

This study, in addition to a similar, albeit single-centre, 2012 study from the same study group based in Salerno, Italy2 represent two of the very few contemporary investigations of relapse seasonality, with the bulk of the precedent literature pre-dating 1990. Indeed, of the 10 population-based studies including at least thirty cases reporting on season of onset of MS exacerbations identified in a 2000 meta-analysis by Jin et al of relapse seasonality, all 10 pre-dated 1985 with some including retrospective data stretching back to 1950.3 Whilst some of the individual studies contributing to this meta-analysis reported no variation in relapse onset, once data was pooled using a weighted mean difference approach, the meta-analysis found a clear seasonal variation of relapse onset, with...
a spring peak and a winter trough, independent of any single-study or site effects. The study further observed a similar seasonality in the timing of MS onset with, again, a spring peak and winter trough. The meta-analysis was however limited by significant between-study heterogeneity, particularly with regards to systematic differences in outcome ascertainment, with only 5 of the 10 studies providing an explicit definition of the primary relapse end-point. Furthermore the meta-analysis included a mixture of prospective and retrospective data. Consistent with the great majority of the analysis on this topic, no relapse data from locations in the southern hemisphere was used in this meta-analysis.

Whilst the Jin meta-analysis was considered to be, methodologically at least, the best available data at the time supporting a predictable, cyclic seasonal gradient in relapse timing, all of the included studies almost entirely pre-date that great real-world moderator of relapse probability in the modern era of MS management – immunomodulatory therapy. Even the very few contemporary studies, exclusively set in or at the very least over-lapping the disease-modifying drug (DMD) era, did not control or otherwise adjust observations of seasonal trends in relapse probability for the influence of treatment. As such it is uncertain exactly how much of the theorised seasonal effect on any underlying biological mechanism promoting relapse in the modern treatment era is moderated or otherwise dampened by exposure to DMDs. In a rapidly changing MS therapeutic environment where DMDs are in part approved for use in preventing or delaying relapse events, it is clear that any statistical model designed to isolate the independent effect of season on relapse timing needs to include the effects, known or suspected, of patient-level promoters or protectors, of relapse risk. Such analytical sophistication requires access to large, longitudinal datasets where patient-level data is routinely and consistently captured. This in part may be a reason for the paucity of appropriately controlled and analysed relapse seasonality studies over the past twenty years.
Outside the 2013 Iuliano study and the 2000 Jin meta-analysis, the remaining literature is largely based on data sourced from single centres, are almost certainly underpowered and present little-to-no formal seasonality testing or modelling. Indeed the majority of reviewed studies rely on crude and entirely unadjusted comparisons of monthly relapse counts or proportions. The trend in the literature favouring presentation of naïve chi-square comparisons over modelling techniques suitable for cyclic, longitudinal phenomena (such as geometric regression or Fourier series analyses) or even generalised omnibus tests of seasonality (e.g. Edward’s, Hewitt’s or Roger’s tests) may have contributed to previous disagreements around whether crude monthly fluctuations in relapse frequency heralded a genuine underlying seasonality or whether this was simply an artefact of reporting. Indeed, this may also have been contributing factor to the effective absence of such studies through the greater part of the late 1980s and 1990s.

One notable and altogether rare exception to this was Kozoil et al’s 2004 use of trigonometric modelling of monthly relapse data, albeit in a very small sample and over a single 12-month observation period only. A cohort of 24 RRMS patients sourced from a United States based RCT of cladribine for RRMS were observed and assessed monthly for 12 months for relapse events. In contrast to much of the precedent studies investigating seasonality in the timing of relapse onset, Kozoil employed a markedly more systematic approach to testing for the presence of seasonality despite the small sample, first plotting the observed monthly relapse data and then formally testing for the presence of cyclic, periodicity across the observed data using a Roger’s variant on the Edward’s test for seasonality. Trigonometric regression was then used to build a model designed to best capture seasonal fluctuation in relapse frequency across the 12 months, notably extending the model to include a random effect to act as proxy for unmeasured patient confounders of relapse probability. Whilst the observed data suggested a late spring/early summer peak with an associated late autumn/early winter trough, the Roger’s test of seasonality was not significant (Rogers R = 0.47, p=0.47). As such there were insufficient evidence from the observed data to support the presence of
a genuine cyclic trend over the 12 observation months, but the analysis, although methodologically appropriate, contained little data and had no power to truly reject or confirm relapse seasonality.

Similarly set in the DMD era and employing formal time-series analyses was a single-centre, retrospective review of 164 relapse events from 96 MS patients managed between 2007 and 2008 at a single clinic in Bologna, Italy.\(^5\) Employing a partial Fourier series (a generalisation of trigonometric regression), the study estimated a bimodal distribution of relapse frequency with a primary or major peak in late-spring, early summer (May-June) and a secondary, smaller amplitude peak in late autumn/early winter (November-December). Fourier-based analysis is a flexible tool for exploring periodic, non-linear distributions, however the modelling of up to four discrete harmonics at once across just 12 months of data from a small sample of 96 patients suggests that this study may have suffered from over-fitting. In particular, over-fitting would increase in severity as a function of the number of trigonometric pairs (i.e. the number of harmonics) included in the model.

Theories around cyclical, periodic patterns in relapse onset as a function of season have been present in the literature over 50 years. A 1959 single centre review of 700 case histories described a July peak (northern hemisphere only) in relapse frequency although this did not represent a statistically significant difference in monthly incidence relative to other months.\(^6\) Seasonality in the timing of MS relapse was later studied several times in the United States throughout the early 1980s. A five-year, single-site prospective analysis of 313 relapse events from 425 probable and definite MS cases in a North Dakota clinic observed a significant seasonality in the timing of relapse onset, with a summer-autumn peak (July-September quarter) associated with a comparatively shallow trough during the October-December quarter.\(^7\) Observed-to-expected relapses by calendar month were compared using simple chi-square tests with no adjustment made for other correlates of relapse probability. Further south in Arizona, a 1983 single-centre, prospective five year case-control study of 178 MS patients and 82 controls also observed that MS relapse events were more frequent in the
autumn and summer months – with peak relapse activity observed over a wide March-August interval regardless of whether the observed data or an alternate 3-month centred moving median smoothing model estimate were used.8 Again, no adjustment was made for potential patient-level confounders of relapse timing. Another small, single centre retrospective study of MS relapses between 1981 and 1985 from 87 patients in Galway, Ireland employed simple correlation analysis to study potential associations between the duration of an relapse event and a variety of local, meteorological factors. The study found a strong and significant, albeit entirely unadjusted, correlation between relapse duration and the total number of sunshine hours.9

The suggestion, however crude methodologically, of a predictable seasonality in the timing of relapse onset, even in very under-powered studies, suggests such a signal may be genuine. A small Danish 1990 retrospective study of 45 MS patients observed seasonal variation on the timing of relapse (p=0.0004) despite its very small sample, with the majority (76%) of exacerbations observed in late winter months.10 Again no formal generalised tests of seasonality were applied and no adjustment for suspected confounders via an appropriate multivariate regression approach was presented, making it difficult to attribute the observed winter peak solely to seasonal influences alone. A more recent, 2005 Spanish study of relapse events in 31 patients retrospectively analysed between 1997 and 2002 applied simple chi-square comparisons of monthly and quarterly relapse incidence, observing a summer peak (June) with a corresponding winter (December) trough (p<0.05) using only a crude chi-square comparison.11

Whilst European and United States based studies dominate the precedent literature, a small number of studies have reported similar observations in non-Western settings. In a Japanese retrospective study of 172 relapse events from 34 MS patients based in Tokyo and Saitama12, simple, entirely unadjusted chi-square analysis was used to compare monthly incidence. Relapse incidence described a bi-phasic distribution with highest incidence during the warmer July-August period with a
secondary peak during January and February ($p$ (chi$^2$)<0.05). The small sample size, lack of formal
test of seasonality and wholly unadjusted analysis of proportions however makes inference and
attribution in this case difficult. Furthermore, the suggestion of a bi-phasic peak, although likely a
product of over-fitting, would lend itself to formal geometric-based modelling techniques which
permits exploration of multiple peak/trough combinations or harmonics. Over the last decade, a
small subset of MS cases has been classified as Neuromyelitis optica, on the basis of its distinct relapse
phenotype and, importantly, a serum auto-antibody marker reacting to the antigen aquaporin-4. 13
Interestingly, a more recent analysis of Japanese patients compared seasonality of MS relapses with
that of neuromyelitis optica (NMO) exacerbations.14 Whilst seasonal variation in was again observed
in MS exacerbations, interestingly no such seasonality was observed for NMO, although this in part
may be secondary to comparative underpowering relative to the MS relapse (131 NMO
exacerbations compared with 295 relapse events).

A retrospective cohort analysis of 420 relapse events from 235 MS patients from Haifa in northern
Israel between 2001-2003 did not observe any suggestion of a periodic relationship between relapse
frequency and either season or month.15 Whilst the primary analysis relied upon a simple analyses of
correlation coefficients, this was supplemented by the derivation of methodologically superior
probabilistic models to simulate expected patterns in distribution of monthly relapse frequency
against which to compare the observed data. The regression-based approach permitted inclusion of
a variety of potential meteorological correlates as adjusting covariates including temperature,
humidity, solar radiation and heat stress index.

Completing the small number of analyses from outside the northern hemisphere, a 2012
retrospective single-centre review of 209 MS patients in the Campinas region of Brazil observed a
summer peak (December-January) and a spring trough (October-November) in relapse incidence.16
The December-January peak coincided with peak UVR and humidity levels. However, like much of
the pre-Kozoil era analyses summarised above, relapse data were analysed using a simple chi-square approach with non-parametric rank-sum tests and no formal test for seasonality, multivariate adjustment or geometric regression analysis was applied.

As previously flagged, the majority of studies reviewed herein attempted no formal cyclic modelling nor any personalised risk adjustment and were all likely to be underpowered, in several cases severely so. Thus whilst the significant majority of studies here consistently report a possible relapse seasonality signal, multiple design, methodological and analytical limitations means that such signals are imprecise at best, both in terms of calendar dates of peak and the exact magnitude of any relapse peak. Many of these studies pre-date the broad availability of statistical analysis packages for programming and running more sophisticated time-series analyses. However, the use of appropriately adjusted and evaluated trigonometric modelling for testing potential seasonal and latitudinal gradients in relapse probability, particularly at a global level, remains a significant gap in the evidence base.

**B.1.2.2 Seasonality and non-relapse MS activity**

A small number of studies satisfying the inclusion criteria (n=3, 8.6%) considered seasonality in non-relapse MS activity or events such as MRI activity or MS-specific admissions or hospitalisations. A 2010 review of thirteen years of MS admissions data sourced from the Scottish Morbidity Register (1997-2009) observed significant peaks in admissions in April and June with troughs in March and October.¹⁷ Unlike much of the literature described above, a formal omnibus test for periodic fluctuation was performed, in the form of a Walter-Elwood test for seasonality which suggested the presence of significant seasonality in the admission data. However only simple, unadjusted chi-square tests of difference between observed and expected admissions were employed in an attempt to quantify the seasonal differences suggested by the Walter-Elwood test. Further, no adjustment was made for the influence of other confounders of admission probability.
A key biomarker for MS activity is the development of focal areas of increase water content on cerebral MRI scans, known as T2 lesions. A 2010 single-location, cohort study of 939 cerebral MRI examinations from 44 untreated relapsing-remitting MS (RRMS) patients in Boston observed a marked spring/summer peak (March-August) in new T2 lesion activity which in turn positively and significantly correlated with region-specific solar radiation data. A mixed effects Poisson-based regression approach was used to model the longitudinal repeated measures MRI data whilst a Monte Carlo simulation was used to assign a random “start” date to the formation of each new lesion (as distinct from taking the date of the scan as a proxy date for lesion “start”), making this one of the more methodologically robust studies of seasonality in MS activity from the available literature, although this study was specific to MRI activity only and not relapse timing.

An earlier 2000 retrospective analysis of 202 monthly MRI examinations from 53 untreated RRMS or secondary progressive MS (SPMS) patients from a south German clinic across 1996 through 1999 observed, specifically, a sinusoidal distribution (single peak and trough) of MRI disease activity over the calendar year, with a clear late spring/early summer peak in the average number of enhancing lesions per month accompanied by an autumnal trough, in both RRMS and SPMS patients. This is consistent with the biphasic seasonal patterns observed for both MS relapse events and onset described above and below respectively. The investigators quite appropriately suggested that statistical models that formally account for seasonal periodicity of, in this example, gadolinium enhancing (Gd+) lesions would make for better isolation of both treatment effects and the influence of seasonality. In other words better separation of the suspected seasonality signal means better isolation of the influence of other prognostic correlates of MRI activity or relapse events.
**B.1.2.3 Seasonality, birth month and MS onset risk**

A large proportion of the reviewed literature specific to seasonality in MS was perhaps the least relevant of the five themes identified and thus will only be covered briefly here. Whilst these references qualified for review on the basis that they dealt with a primary theme of the interaction of seasonality and MS risk, these studies dealt primarily with the study question of a potential correlation between month and/or season of birth with later risk of MS onset. This hypothesis has been examined often, based on the theory that early life environmental exposures, including those in-utero, alter immune system development especially the powerful negative thymic selection of autoimmune reactive lymphocytes. The majority of references categorised under this theme used aggregate population data sourced from such birth registries and administrative databases. As such, whilst many of the sample sizes presented in these studies are indeed impressive, patient-level data is not available for most and thus analyses of birth season-MS risk correlations have largely not been adjusted for patient-level modulators of MS risk. Consistent with the data around seasonal patterns in relapse or MRI activity, the vast majority of data is again sourced from northern hemisphere locations.

Most notably within this pool of reports were two systematic reviews from 2012. A systematic review of 15 published studies and two congress abstracts analysing birth month or season observed a dominant spring peak in birth month in subjects who subsequently went on to be diagnosed with MS with an autumnal trough. A similar pattern was observed in the southern hemisphere, although this was based on a single site only, with, again, a spring peak (November) and autumn (April) trough. Importantly the investigators observed that this seasonal pattern was most pronounced in areas of higher latitudes away from the equator and almost non-existent in locations with a higher solar index (typically locations more proximal to the equator).
The second review was the more methodologically robust. A systematic review and meta-analysis from the same year combined month of birth data from 151,978 MS patients from 10 studies. Once the data was pooled and assessed for significant inter-study heterogeneity (which ranged from 0-91%), linear regression was used to first test the correlation between month of birth and subsequent MS onset (quantified as the ratio of observed to expected for each calendar month). Peak observed-to-expected ratio (O:E) was observed in April (ratio=1.05, p=0.05) with a corresponding autumn trough (November ratio 0.92, p=0.04). Only December demonstrated a significant association between ratio and latitude (p=0.039). Of note, the meta-analysis was limited to data from northern hemispheric locations only, spanning a latitude range of just over 30 degrees in total (from 32.3 degrees N (Tel Hashomer, Israel) up to 62.4 degrees N (Helsinki, Finland)). The meta-analytical model chosen, a simple linear regression, also presumes, as the name suggests, linearity in the interaction between latitude and O:E ratio and no non-linear alternatives were trialled.

Of note from the remaining studies reviewed from this category, a 2013 retrospective study of 421 MS patients sourced form a single Portuguese (MS) clinic combined a Hewitt’s test of seasonality with a non-parametric rank-sum approach to test for seasonality in month of birth in MS patients relative to a healthy control group extracted from the Portuguese National Statistics Institute births registry for the same geographical region. Whilst the six-month interval spanning the birth months of July through December correlated with higher MS incidence, this was not significant on the Hewitt test. No adjustment was made for patient level confounders. Indeed, many of these month-of-birth studies have been criticised for inadequate control of confounding. This is in part secondary to a reliance on aggregate, population level data for which patient-level factors are not available as described above. However many of these studies using population averaged statistics also use an erroneous assumption of homogeneity in the distribution of other potential prognostic correlates of MS onset across patients from which these aggregate summary measures are sourced. This is akin to assuming that the risk profile for MS onset or activity for all patients contributing to a population-
Timothy Denis Spelman (58172)

averaged metric is effectively the same, which is unlikely. This underscores the importance of using appropriate data (i.e. patient-level) and an appropriate, flexible methodology (e.g. regression-based techniques permitting covariate adjustment) to minimise bias and increase the chances of isolating a genuine, independent seasonal and/or latitudinal effect.

B.1.2.4 Latitude and MS activity

Ten (28.6%) of the reviewed references dealt with a primary study question of a potential correlation between MS activity and latitude. All the reports explored latitudinal gradients in MS onset only. Not a single one of the included studies investigated a latitude signal in, explicitly, the timing of relapse onset.

Previous epidemiological studies and reviews have observed an often marked latitudinal gradient in the incidence and prevalence of MS with polar latitudes away from the equator correlating with greater frequency of MS.\textsuperscript{25-28} Such a gradient in MS onset has been broadly observed in locations across both northern and southern hemispheres including in Europe, North America, Japan and Australia.\textsuperscript{29-32} However previous observations of inconsistencies in the latitudinal gradient hypothesis, particularly within Mediterranean Europe and Northern Scandinavia where no such signal had been observed, prompted several investigators to suggest that such gradients may be a statistical, economic or migration artefact rather than an authentic epidemiological latitudinal-modulated phenomenon.\textsuperscript{27,28} Furthermore, other studies have suggested this latitudinal gradient in prevalence at least may in fact be in “temporal retreat”, most notably a 2008 systematic review of 27 MS prevalence studies reporting age and sex-stratified incidence rates for MS.\textsuperscript{33} Whilst studies set in higher latitudes associated with a higher incidence of MS relative to lower, more equatorial latitudes (incidence increasing by between 30\% and 50\% in females and males respectively for every 10 degrees of latitude away from the equator), this gradient was steeper prior to 1980 compared with post-1980. Notably an increase in MS incidence at lower latitudes after 1980 was identified as
a primary driver of this blunting of a previously well-delineated and statistically significant latitude gradient. However the data extracted from the eligible studies was not formally pooled and no measure of inter-study heterogeneity was provided. Thus systematic differences in study methodology, end-point definition and patient characteristics are likely significant confounders of this observed temporal dampening in the latitude gradient. In particular, the availability of injectable DMD therapy post-1980 would appear a key potential confounder, not controlled for in this analysis. If DMD treatment at clinically isolated syndrome reduces the rate of conversion to clinically definite MS and if such treatment is used inconsistently across latitude (e.g. used more within more polar latitudes) then this may potentially mask a latitudinal gradient.

More recently, Koch-Henriksen et al’s 2010 retrospective review and formal meta-regression of 178 MS epidemiological studies dating back to 1965 confirmed the latitudinal incidence gradient for Australia and New Zealand but found no evidence for it in the northern hemisphere. A subsequent, more comprehensive 2011 meta-analysis by Simpson Jr et al of 650 MS prevalence estimates sourced from 321 peer-reviewed studies however found a clear and significant global increase in prevalence per degree of latitude away from the equator, a signal evident in both hemispheres. It should be noted however that the majority of the studies included in the Koch-Henriksen and Simpson Jr reviews pre-dated the DMD treatment era, a suspected correlate of both conversion to clinically definite MS and relapse probability.

Consistent with the literature of relapse seasonality, the vast majority of onset studies were again sourced from either a European or US setting. Of the very limited number of studies sourced from alternate settings, a 2011 systematic review focusing on MS prevalence studies from Latin and South America identified 10 studies across six countries. Using linear regression to model the association of latitude with reported prevalence, the investigators estimated a significant latitudinal gradient from Panama through to Argentina, with every additional degree of latitude significantly associated
with a 0.33 per 100,000 increase in MS prevalence. This was supported by a strong, significant Pearson’s correlation coefficient ($r^2 = 0.8$, $p<0.001$).

### B.1.2.5 Seasonality and ultraviolet radiation

The final review theme captured a small (2, 5.7%) number of studies that attempted to link observations of seasonal fluctuations in MS onset or activity to a underlying UVR/Vitamin D regulated mechanism of action by including an analysis of solar radiation and, where available, vitamin D levels. A matched case-control study of 29,994 patients in England and Scotland combined health registry and administrative database data to explore seasonality in the risk of multiple sclerosis based on season and month of birth and the timing of pregnancy trimester.\(^{36}\) Using a generalised Poisson regression approach fitting a single sine and cosine pair to generate a sinusoidal model to capture seasonality birth month data, monthly MS risk was observed to correlate inversely with predicted UVR levels during the second trimester of pregnancy (Spearman’s rho = -0.499, $p<0.001$) whilst serum 25(OH)D levels during both the second and third trimesters were similarly estimated to inversely correlate with MS risk (rho = -0.49, $p<0.001$ and -0.44, $p<0.001$ respectively).

A 2010 cohort study of 967 MS patients sourced from the US Veterans Health Administration MS Surveillance registry mapped place of birth to annual mean daily solar radiation data sourced from the US Weather Bureau.\(^{37}\) The study observed MS patients born in winter months and in a geographical location classified as “low solar radiation” experienced onset of first symptoms an average of 2.8 years earlier than MS patients born in non-winter months and in non-low solar regions ($p=0.02$). The authors relied on adjusted, multivariate linear modelling of age of onset to separate out birth season effects from birth place solar radiation classification level. However the preference for, in this case, a parametric model of the mean presumed age at MS onset was normally distributed. Whilst general reference is made to testing variable distribution in the
methods, no formal test of skew was presented to confirm that a parametric model was the most appropriate choice for the analysis.

**B.1.3 Conclusion**

The state of the current literature is that there is little individual analysis of sufficient power using appropriately robust statistical methodology to address the direct question of whether the suspected seasonality in the timing of relapse onset is latitude dependent, but the overall impression of seasonality across all studies is strong. Furthermore, outside systematic reviews and meta-analyses of separate studies, limited by considerable between-study heterogeneity, no single-source study has been attempted to address these questions at a truly global level. In light of previous disagreements around the authenticity of a previously observed seasonal and/or latitudinal signal, a large dataset such as MSBase with its broad geographical coverage, a generous observation period, unified data entry and definition protocol and, critically, patient-level data to adjust for multiple potential confounders – particularly as demographic, relapse onset and treatment exposure data is collected in the same minimum dataset, is ideally placed to extend the search for independent environmental correlates of relapse risk.

**B.1.4 References**


B.2 Hypothesis 2: Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis

B.2.1 Search Strategy

References for this review were identified using a search of the PubMed and Scopus electronic databases of peer-reviewed literature using a pre-specified review protocol. The search included the terms “multiple sclerosis” and “conversion” in combination with “clinically isolated syndrome” or “clinically definite multiple sclerosis” within either the article title or abstract published between 2000 and 2013. The references of evaluated articles were then manually screened for additional publications that satisfied the search criteria. The Cochrane review library was then reviewed to identify relevant systematic reviews and meta-analyses not identified in the initial PubMed/Scopus search. The search included references in both English and languages other than English. Case reports were not reviewed. After removing duplicate records, the full reports of each reference satisfying the search criteria were reviewed for relevance and quality to arrive at the final reference set.

B.2.2 Results

The initial screening consisting of the electronic database search and manual review of references identified 62 references. A total of 16 publications (25.8% of the initial screen) were rejected, 12 on the basis of insufficient relevance and/or quality and 4 as duplicates. The PRISMA flow diagram in Appendix I.2 summarises the screen and selection process. Of the 46 papers satisfying the inclusion criteria, two broad analysis themes were identified, 1) studies of a single candidate predictors of conversion and 2) studies analysing the performance of multiple, concurrent prognostic correlates of subsequent conversion to CDMS. The former were generally limited to exploratory studies of potential biomarkers or MRI metrics.
B.2.2.1 MRI

Of all the candidate demographic, clinical and examination predictors of conversion to CDMS, the most frequently studied of the eligible references were MRI features at time of CIS. This is in part a reflection of the gradual incorporation of MRI lesion metrics into the various iterations of formal diagnostic criteria used for MS diagnosis including the Barkhof-Tintore criteria\textsuperscript{1} and the more recent Swanton’s modifications.\textsuperscript{2} The presence of any asymptomatic typical lesions on cerebral MRI at time for CIS is strongly predictive of clinical conversion from CIS to RRMS, ie relapse onset. A 2013 meta-analysis of 23 cohort studies from 1991 through 2010 capturing 1122 CDMS conversion events from 1827 CIS patients observed the presence of any lesions on cerebral MRI was associated with 3.71 times the risk of conversion (pooled Relative Risk (RR) 3.71; 95% CI 3.27, 4.21).\textsuperscript{3} The risk of converting was higher in what was categorised as a “medium” burden group (defined as between 4-9 T2 hyperintensive lesions) relative to the “low” burden group (defined as 1-3 lesions) with the low burden group associated with a 34% relative reduction in conversion risk (RR 0.66; 95% CI 0.45, 0.95). Whilst appropriate heterogeneity statistics were reported, funnel plots cited in the methods as a device for quantifying inter-study bias were not presented. The studies included in the meta-analysis varied markedly with regards to follow-up, definitions, ascertainment frequency and length of observation period. In perhaps the most significant limitation, the authors erroneously interpreted the Hazard Ratio point estimates reported by the majority of studies included in the meta-analysis to be equivalent to, and thus exchangeable with, relative risk in deriving the key results as pooled RRs. This is incorrect as the former is a metric of instantaneous risk observed prospectively over the study period whilst the latter is a cumulative metric only calculable at the end of a pre-defined observation period, and thus these are not interchangeable.

Examining non-lesional metrics, a 2013 prospective study of serial MRI examinations in 216 CIS patients over two years observed both thalamic and lateral ventricle volumes were both correlate
with conversion risk.\textsuperscript{4} Using multivariate, mixed effects regression to model temporal changes in MRI metrics, Zivadinov et al demonstrated that an decrease in thalamic volume over the observation period (captured as % change) was associated with a 15% decrease in the odds of conversion (OR 0.85; 95% CI 0.77, 0.95). Conversely every percentage increase in lateral ventricle volume (reflecting loss of brain tissue causing CSF-filled ventricular system ex vacuo expansion) was correlated with 1.07 times the odds of CDMS development (OR 1.07, 95% CI 1.04, 1.1), adjusting for age, time from CIS to baseline and change in treatment status. Whilst the multivariate approach permitted some control for known or suspected confounders of conversion including treatment exposure, these models were not adjusted for other potential confounders including neuroanatomical location of first symptoms, sex, baseline EDSS, OCB on CSF examination or spinal MRI confounders. Whilst serial MRI assessment was prospective, this was not a strictly “seen-from-onset” (seen from CIS) cohort and whilst some attempt to control for this was attempted at the level of the multivariate regression (through inclusion of time from onset to baseline as a fixed effect in the multivariable model), residual bias may influence the observed results, particularly in that subset of patients for which a truly “baseline” MRI (i.e. at CIS) was available.

A 2012 longitudinal, exploratory cohort analysis of 217 CIS patients from the prospective, multi-centre SET study (Study of Early InterferonB-1a Treatment in High Risk Subjects after CIS) examined the predictive value of volumetric MRI metrics for subsequent conversion.\textsuperscript{5} Of the volume and area markers studied, 6-month change in corpus callosum (CC) mid-saggital area and overall volume of T2 lesions at baseline were both associated with increased likelihood of 2-year conversion under a multivariate logistic approach (logit $\beta$ coefficient 0.4, $p=0.008$; $\beta= 0.6$, $p=0.02$ for CC and T2 predictors respectively). The relative frequency of conversion events, however (n=92, 42%), suggests that these logit point estimates may over-estimate the true risk of conversion, as log-odds or odds ratios best estimate the relative event risk where the study end-point is sufficiently rare (<20%). Whilst those MRI metrics identified as risk factors on logistic regression were likely to slightly over-
estimate the true risk of conversion, these MRI risk factors were then synthesised in a time-to-event analysis to estimate the rate of conversion associated with patients satisfying an increasing number of these volumetric predictors. Consistent with the results of the logistic modelling, the greater the cumulative number of positive MRI predictors reported, the greater the conversion hazard. Two MRI predictors associated with 6.5 times the rate of conversion (HR 6.5; 95% CI 3.4, 12) compared with a smaller, albeit non-significant rate when only 1 MRI predictor was present (HR 1.2, 95% CI 0.7, 2.2). The cumulative hazard described by these various levels of MRI predictor were not equivalent to a cumulative risk of conversion, as claimed by the report.

Swanton’s own 2010 study of the utility of early MRI in 142 optic neuritis patients in conversion risk quantification isolated both periventricular (aHR 3.50; 95% CI 1.67, 7.33) and presence of Gd lesions (aHR 2.09; 95% CI 1.21, 3.61) as independent predictors of conversion, in addition to sex, using a multivariate Cox proportional hazards (PH) model. A 2008 analysis of 15 years of prospective follow-up in a multi-centre cohort study of 389 patients initially presenting with optic neuritis reported an incremental increase in conversion hazard with an increasing number of brain MRI lesions, ranging from 2.8 times the rate of conversion in the presence of a single lesion (HR 2.8; 95% CI 1.7, 4.7; reference = zero lesions) to 4.5 times the rate with multiple lesions (HR 4.5; 95% CI 3.0, 6.6). It is unclear, however, whether these estimates were adjusted in a multivariate Cox model or whether a test for hazard proportionality was conducted to confirm the appropriateness of the model choice.

Post hoc analyses of MRI data sourced from pivotal DMD trials in CIS has also provided high-quality data on MRI imaging metrics as prognostic correlates of conversion, although these tend to be limited to relatively short study observation periods. A 2009 analysis of 197 conversion events from 468 CIS patients from the three-year BENEFIT core plus extension trial (early IFNβ vs delayed) observed that the strongest associations with conversion over 36 months were observed with a
minimum 9 T2 lesions on baseline MRI (aHR 1.64; 95% CI 1.15, 2.33) and a minimum three periventricular lesions (aHR 1.66; 95% CI 1.14, 2.41), adjusting for age, sex, treatment intervention, steroid use, symptom presentation type and disease onset type. There was no difference in any MRI criterion tested by treatment intervention (early vs late IFNβ, p>=0.2 for all). Interactions between concurrent explanatory variables in the Cox PH model were tested. However a test for hazard proportionality was not reported.

The incorporation of lesion metrics in diagnostic criteria also provides a useful tool for studying the accuracy and discrimination of such metrics, particularly as measured at the time of CIS presentation, for predicting subsequent conversion to CDMS. Lo et al retrospectively reviewed baseline MRIs of 64 CIS patients between 2001 and 2006 to compare the performance of, specifically, the revised McDonald MRI criteria relative to Swanton’s in predicting conversion to CDMS. Whilst accuracy, sensitivity and negative predictive value were higher under Swanton relative to McDonald (accuracy: 81% vs 78%; sensitivity: 60% vs 53%; NPV 74% vs 71%), none of these differences were statistically significant. A sample size calculation or power analysis suitable for ROC-type area-under-the-curve analyses, however, was not provided, thus it is unclear whether these differences are genuinely non-significant or authentic, albeit underpowered, differences. A much larger sample would thus be required to explore this further. A similar sized 2010 Spanish single-centre study of 79 CIS patients with baseline MRI available compared the accuracy of Barkhof-Tintore’s criteria with the simpler Swanton’s criteria as previously described in the literature. Consistent with the Lo study, for the 75.7% of the sample who converted under Swanton’s returned superior sensitivity relative to Barkhof-Tintore (91.2% vs 71.9%) but a marginally inferior specificity (68.1% vs 77.2%), translating into a superior accuracy under Swantons (point estimate 84.8%; 95% CI 75.0, 91.9%) compared to Barkhof-Tintore (73.4%; 95% CI 62.3, 82.7%). A supplementary Kaplan-Meier time-to-conversion analysis observed a greater hazard of conversion in those CIS patients satisfying Swanton’s relative to CIS patients not satisfying Swanton’s (HR reported as “close to 6”;

50
95% CI 2.00, 5.97) than the same comparison using Barkhof-Tintore (HR not reported; 95% CI 2.00, 5.97; reference = not fulfilling Barkhof-Tintore). However when time to conversion under Swanton was directly compared to Barkhof-Tintore, no such difference was observed (p(log-rank)=0.44). As was the case with the Lo study, the small sample size studied and absence of a formal power analysis suggests this analysis was likely sub-optimally powered. Similarly, a 2004 long-term, albeit small, follow-up cohort study of 26 conversion events in 42 CIS patients (median follow-up 8.7 years) localised 2+ infratentorial lesions on baseline MRI as the strongest predictor of conversion on Cox PH modelling, where predictor “strength” was defined on the largest likelihood ratio chi-square value (11.27) in the model.¹¹

Limited data is available from outside western settings with only a single study identified by this review. A very small Brazilian cohort study of just 12 CIS patients reported correlations between, specifically, infratentorial and spinal cord lesions with time to conversion.¹² Whilst a Cox regression model was cited in the methods as the analytical tool, no HRs were presented and the specific type of Cox model was not stated nor any tests of fit or proportionality provided. Thus these results can at best be considered descriptive only.

**B.2.2.2 Disease-modifying drugs**

Perhaps not surprisingly, given their subsequent involvement in DMD regulatory and approval processes, the influence of DMD treatment on reducing the risk of converting to CDMS in CIS patients is also a frequently studied area. However the existing evidence base does suffer from similar limitations as the data supporting MRI correlates of conversion risk, primarily those of under-powering, inadequate-to-absent methodological control for confounding or selection bias and mixed generalisability, particularly to real-world populations. At the highest level of evidence, randomised clinical trials have provided the strongest evidence supporting the efficacy of DMD treatment in delaying or preventing conversion to CDMS. The DMD’s IFNβ-1a, IFNβ-1b and glatiramer acetate
have been trialled in high-risk CIS patients compared to placebo, and all have been shown to significantly reduce the proportion of patients converting to CDMS.\textsuperscript{13-18}

Extension studies and exploratory post-hoc analyses of pivotal clinical trial data also provide relatively high-quality comparisons of, typically, the efficacy of early treatment (the originally randomised treatment arm) in delaying or decreasing the risk of conversion relative to late or delayed therapy (the original placebo arm switched to active treatment subsequent to the original trial phase). The open-label phase of the multi-centre PreCISe trial of glatiramer acetate in CIS observed early GLA (n=198 from the original randomised treatment arm) was associated with a 41% reduction in conversion to CDMS relative to delayed initiation of glatiramer (n=212 form the original placebo trial arm) (HR 0.59; 95% CI 0.44, 0.80).\textsuperscript{19} Whilst the non-blinded end-point analysis was entirely exploratory (in that they were not explicitly powered for as part of the original randomised PreCISe design) the effect sizes were nonetheless comparatively large and confounders generally remained well balanced between the early vs late treatment groups. A targeted sub-group analysis looking for differential responders delineated by demographic and clinical disease factors confirmed that the efficacy advantage favouring early glatiramer was largely consistent across sub-groups, although some of these sub-groups were under-represented and thus likely underpowered (e.g. males, <9 T2 hyperintensive lesions).

Consistent with the PreCISe extension, the 5-year extension study of the IFNβ BENEFIT trial observed a 37% reduction in the rate of conversion to CDMS in the early IFNβ treatment arm relative to the delayed treatment group (HR 0.63; 95% CI 0.48, 0.83).\textsuperscript{20} Whilst this involved a breaking of the original randomisation (80% of the originally randomised treatment (intervention) arm and 70% of the placebo group were enumerated in the 5-year (extension) study), the dose difference blinding was preserved into the extension phase. The Cox PH model used for the primary analysis was
adjusted for age, sex, MRI parameters, steroid use and focal nature of onset (mono vs multi) to adjust for imbalance introduced through the breaking of randomisation.

Similar to the literature around MRI correlates of conversion, only a single study sourced from a non-western setting specific to DMD treatment as a conversion risk mitigation strategy was identified. A 2007 Iranian clinical trial randomising 104 CIS patients to IFNβ-1a and 98 to placebo observed a significantly smaller proportion of treated patients converted to CDMS relative to the placebo arm (36.6% vs 58.2%, p(rank-sum)<0.003). Whilst the intervention and placebo arms were well balanced at baseline for observed characteristics, a standard time-to-event analysis of the conversion end-point was not reported and thus the crude effect size is entirely naïve to perhaps the more important question around the temporality of conversion –i.e. how fast or how much delayed is conferred by treatment.

**B.2.2.3 Cerebrospinal Fluid examination**

Amongst the battery of tests and examinations typically undertaken when a patient first presents with a potential clinically isolated syndrome, examination of the cerebrospinal fluid has become an increasingly popular study target as a potential prognostic correlate of later conversion, particularly with regards to the presence of oligoclonal banding (OCB). A systematic review and meta-analysis of 21 studies reporting OCB positivity across 2685 CIS patients used an inverse variance meta-regression to estimate that CIS patients with OCB-positive CSF examination at CIS were associated with 9.88 the pooled odds (HR 9.88; 95% CI 5.44, 17.94) of conversion relative to OCB negative CIS. Between-study heterogeneity was reported and the meta-regression logistic model extended to include a random effect in the presence of moderate to high heterogeneity. In addition, study bias was further assessed through funnel plots making for a methodologically robust report.
A 2013 retrospective review of 205 patients from a single centre in Modena, Italy presenting with CIS between 1993 and 2008 for which a baseline CSF was available, observed a correlation between the pattern of oligoclonal banding on baseline CIS and subsequent rate of conversion to CDMS. Specifically Ferrato et al used a Cox time-to-event model to estimate that, on average, the presence of CSF-restricted IgM oligoclonal bands was weakly associated with 1.5 times the rate of conversion (unadjusted HR 1.5; 95% CI 1.00, 1.22). This association was not adjusted for other likely prognostic correlates of conversion in a multivariate model and given the weakness of the univariate correlation it is likely such an association would drop out of significance on a fully adjusted model. Furthermore the exact variety of Cox model (e.g. proportional hazards, Weibull, Gompertz) was not stated in the report, making it difficult to assess whether the presented model complied with the relevant underlying assumptions (e.g. proportional hazards, parametric distribution of hazards).

Several studies have extended the accuracy analyses of baseline lesion metrics on MRI at CIS to further test the relative accuracy of the presence of OCB for subsequent conversion. A 2006 single-centre Spanish discrimination study of CSF samples from 52 CIS patients was one of the earliest to observe a correlation between the presence of oligoclonal IgG bands on CSF examination at or around the time of CIS with the probability of subsequent conversion to CDMS. Specifically the presence of OCB was associated with a comparatively high accuracy (91.4% sensitivity, 94.1% specificity) for conversion. This out-performed MRI (74.2%, 88.2%). Although based on a smaller sample, these results represented a marked improvement on Tintore’s earlier 2001 analysis (n=112 CIS) that reported an OCB sensitivity of 81% and specificity of just 43% for conversion, translating into a modest accuracy of 52%. Notably, a separate, later and larger Tintore prospective study of 572 CIS patients reported that conversion was both frequent (415, 73%) and that a OCB positive CSF at baseline (CIS) was associated with 1.7 times the rate of conversion (aHR 1.7; 95% CI 1.1, 2.7), adjusting for baseline MRI.
B.2.2.4 Biomarkers

The poorest quality evidence tends to cluster around a number of entirely exploratory analyses for novel, candidate biomarkers of conversion risk.\textsuperscript{27-33} This is largely secondary to the typically very small sample sizes studied which in part reflects the likely expense and logistic difficulties in both accessing and processing samples, at least relative to the traditional clinical and radiological examinations that form the core of the battery of tests for when a patient presents for a CIS work up. Of the more notable studies, a 2013 single-centre, prospective cohort study of 249 CIS patients enrolled in the Barcelona cohort between 1995 and 2007 and for which serum samples were available, studied the performance of the gMS-Classifier\textsuperscript{2} predictive algorithm on inclusion of the presence of the anti-glycan P63 IgM antibody in predicting subsequent time to conversion.\textsuperscript{27} A classifier-positive status (75 or 30.1\% of the sample) was associated with 1.8 times the rate of conversion relative to classifier-negative patients on multivariate Cox proportional hazard regression (HR 1.8; 95\% CI 1.1, 2.8), adjusting for, alternately, Barkhof criteria and CSF OCB or the presence of T2 hyperintensive lesions on baseline MRI and OCB. Arrambide et al further supplemented the Cox modelling with a comparative ROC area-under-the-curve discrimination analysis which demonstrated that the addition of the gMS-Classifier\textsuperscript{2} metric to a multivariate, predictive model was associated with a significant gain in AUC including either Barkhof (delta AUC 0.0415, p=0.012) or T2 lesion count (delta AUC 0.0467, p=0.009). Whilst a Hosmer & Lemeshow goodness-of-fit test was used as a measure of calibration for comparing competing model structures, hazard proportionality for the multivariate Cox models was not reported. Whilst the characterisation in the report of gMS-Classifier\textsuperscript{2} as a nominally “independent” predictor of subsequent conversion is valid with regards to those explanatory variables, it was combined within the model (Barkhof, T2 lesion count), and was not demonstrated to be independent of those key prognostic correlates of conversion risk not included in the modelling (e.g. treatment exposure, age, sex, baseline EDSS, first symptoms location). Conversely these results did not corroborate an earlier 2012 analysis of serum samples taken from 258 CIS patients in the BENEFIT study (61\% of the original randomised sample) which
failed to support gMS-Classifier I as an independent predictor of conversion to CDMS, although it was associated with an increased rate of confirmed disability progression on log-rank analysis (p=0.012).\textsuperscript{30}

Controversially, a small, prospective, Swiss 12 month cohort study of anti-myelin antibodies in 39 CIS patients observed a significantly higher rate of conversion within the antibody positive group compared with an antibody negative result (p(log-rank)=0.01).\textsuperscript{33} Consistent with many of the candidate biomarker studies reviewed, no adjustment was made to control for confounding nor any sample size calculation or power analysis provided. Thus this result can be considered at best exploratory and unverified. A similar retrospective cytology analysis of CSF samples from a cohort of 58 CIS patients taken between 2006 and 2010 observed a significantly higher median baseline CSF white blood cell count (WBC) in patients who subsequently converted to CDMS (median 5 WBC/mm\textsuperscript{3}, range 0-45) when compared to non-convertors (median 2, range 1-12, p=0.0034).\textsuperscript{29} Simple, univariate tests of difference were used in the analysis and this result was not adjusted for any confounder, known or suspected. In addition, no a priori sample size calculation or power analysis was provided, nor any clinically meaningful difference cut-off identified or suggested. Finally, a 2011 prospective single-centre study of 55 consecutively enrolled CIS patients and 40 healthy controls observed a high CD5\textsuperscript{+} B cell count was associated on multivariate Cox regression with 4.3 times the rate of conversion (aHR 4.3; 95% CI 1.9, 9.5).\textsuperscript{32} However the wide confidence interval around the HR point estimate suggest a rather imprecise average effect. The report did not state which adjusting covariates had been included in the multivariate model and the observed results have, to date, yet to be independently confirmed. Ultimately, all of these studies suffer from likely under-powering. The largest, most methodologically robust analysis available on the topic, and analysis of 462 patients, did not observe an association between anti-myelin antibodies and progression.\textsuperscript{34}
B.2.2.5 Developmental and environmental correlates

A small subset of the reviewed studies focused on potential non-clinical predictors of conversion risk. Two such studies analysed smoking as a primary risk factor. A 2013 single-centre Turkish cohort study of 95 CIS patients comparing smokers to non-smokers at time of CIS observed no difference in the rate of conversion by smoking status using multivariate Cox PH regression (HR 0.84; 95% CI 0.47, 1.50) adjusted for age, sex, T2 lesion count and white matter lesions on baseline MRI and EDSS. While the study was a priori powered to detect a minimum effect size of 0.243 based on the original sample of 134, no adjustment or inflation was made for drop-out, eventually observed at a moderate 26.4% of the original sample size, and thus the sample size was likely underpowered.

Furthermore the Cox model presented was not tested for interactions and a test for hazard proportionality was not presented. A 2008 Austrian observational cohort study of 129 CIS patients with disseminated white matter lesions and positive OCB in baseline CSF, followed up for 3 years observed smokers converted to CDMS at 1.8 times the rate relative to non-smokers (aHR 1.8; 95% CI 1.2, 2.8). Whilst the study adjusted smoking status at CIS for age, sex, symptom location, T2 lesion count and IFNβ treatment, a test of hazard proportionality for the final multivariate model was not reported and, similar, to many of the studies summarised herein, hazard ratios were erroneously presumed to be directly exchangeable for the relative risk - a frequent mistake observed across the precedent literature.

A single study explored cognitive impairment at CIS presentation as an additional candidate predictor of conversion to CDMS. This 2010 single-centre Italian observation study employed both the Brief Repeatable Neuropsychological Battery (BRB) and the Stroop Tests to assess a cohort of 61 consecutively enrolled CIS patients for cognitive impairment at baseline (CIS). Cox regression was then applied to estimate the independent effect of cognitive impairment on conversion probability, adjusting for age, gender, education level, disease duration, onset symptomatology, baseline EDSS, McDonald’s dissemination in space (DIS) criteria, OCB presence on baseline CSF examination, fatigue...
and depression. On modelling, the failure of a minimum three tests of cognitive impairment was associated with 3.3 times the rate of conversion (aHR 3.3; 95% CI 1.4, 8.1). However, given only 26 (46%) conversion events were observed in the study it is probable that a multivariate model adjusting for 10 confounding explanatory variables was markedly over-fitted, even under then most generous interpretation of the number of events permitted per explanatory covariate included in the model. Thus the observed effect sizes need to be interpreted with caution and would, as a minimum, require replication in a much larger sample. Hazard proportionality was once again not reported.

B.2.2.6 Methods for isolating independent predictors of conversion

A 2013 prospective observational cohort study of 168 CIS patients sourced from the GERONIMUS multicentre Italian study examining four years of post-CIS follow-up employed a standard multivariate approach to identifying independent predictors of conversion. Such a multivariate Cox PH regression modelling was used to isolate independent demographic, clinical and examination correlates at CIS of increased conversion probability. Younger age (<=32 years), involvement of two or more functional systems, positive Barkhof criteria at baseline MRI and absence of DMD treatment prior to conversion were all associated with an increased rate of conversion. Interaction testing was not reported, nor was a test of hazard proportionality.

Whilst adjusted modelling has been the favoured approach for studying multiple potential predictors of conversion concurrently, a small number of studies have attempted alternate approaches to isolating predictor signals. Martinelli et al explored an integrative approach to synthesising multiple risk factors at CIS into a single standardised metric of conversion risk intended for personalised risk assessment. This single-centre cohort study of 227 CIS patients used multivariate logistic regression to first identify risk factors for conversion at 6, 12 and 24 months respectively, observing that the presence of OCB in baseline CSF was associated with increased odds of both 12-month (aOR 3.1;
95% CI 1.04, 7.98) and 24-month conversion (aOR 3.57; 95% CI 1.27, 11.98). Multimodal evoked potentials registered in the 4th quartile of conduction delay were similarly correlated with an increased risk of early 6-month conversion (aOR 1.74; 95% CI 1.49, 15.93). This compared with the overall model across the full observation period where only the number of T2 lesions on baseline cerebral MRI, whilst not associated with conversion risk in either the 6, 12 or 24 month models, was associated with the odds of any observed conversion (>9 T2 lesions aOR 5.7 (95% CI 2.09, 12.77). Whilst these results suggest that MRI, CSF examination and evoked potentials may be good candidates for the derivation of a combined personalised conversion risk score, no such score was attempted or presented in this study. Furthermore the wide confidence intervals around the point estimates and a pattern of borderline, as distinct from strongly, significant predictor signals suggest a larger sample with more conversion events would be required to return estimates of effect of sufficient precision to confer appropriate discrimination on any personalised risk score. Finally the use of odds ratios over hazard ratios as a proxy for risk in this example may have its limitations. Whilst odds ratios can provide robust estimates of the relative risk where the conversion outcome is sufficiently rare, conversion in this study was relatively frequent with 120 (52.9%) of the sample converting to CDMS within the observation period, consistent with other studies of conversion in real-world settings. It is therefore likely that the odds ratios presented in this study over-estimate the true relative risk. In addition the odds ratios ignore the temporal dimension, capturing only if an event happens, not the time to or rate of conversion. As such a time-to-event Cox –based model using HRs to capture effect size may have been preferable in this case.

A 2013 study eschewed traditional exploratory multivariate regression in favour of support vector pattern recognition machine learning algorithms to isolate combinations of patient and disease factors with a particular focus on MRI factors which optimised early identification of those CIS patients most likely to progress to CDMS. Using baseline data from 73 patients presenting to a single clinic in London, all permutations and combinations of available baseline factors were
assessed for predictive accuracy. Maximal accuracy (86% correctly classified) and positive predictive value (PPV) was achieved under a combination of age, sex and location of onset symptoms coupled with either; 1) baseline MRI lesion count and brain volume or 2) lesion load, mean lesion size, minimum and maximum lesion size and gray matter volume. Whilst the accuracy achieved here may approach clinical usability, the results are limited by the small sample size and lack of validation, internal or external. The single centre design of the study further limits generalisability. This extended an earlier analysis by the same study group employed a similar methodology in a smaller cohort (n=44). A comparable, albeit smaller, range of risk factors (demography, disease activity features and MRI metrics) was studied. Optimal univariate accuracy was achieved using lesion count, lesion load and average lesion proton density intensity (accuracy=0.614) used in isolation. Age at CIS was associated with the poorest accuracy (0.409). When combined into a multivariate model, accuracy was maximised in a model combining lesion load, average lesion T2 intensity, lesion with shortest horizontal distance from the centre and neuroanatomical first symptoms location (accuracy=0.773, sensitivity=0.773, specificity=0.773). Both these studies employing support vector pattern recognition algorithms support what has long been suspected clinically- that a combination of risk factors at CIS will outperform any one, single factor when predicting subsequent conversion. This, intuitively at least, makes good sense in that conversion, as a mechanism, is difficult to predict in clinical practice – a feature often symptomatic of an underlying complex and multifactorial mechanism.

Beyond the frequentist approaches to statistical analyses employed by the reviewed literature, Bayesian statistics have also been used, albeit rarely, to localise subsets of maximally discriminant predictors of conversion. A 2010 retrospective chart review of 116 ON/CIS patients presenting to a single Canadian neuro-ophthalmologist between 1990 and 1998 adopted a Bayesian approach to estimate the sensitivity and specificity of various MRI criteria in predicting conversion. The observed accuracy was maximised under at least 1 T2 lesion (sensitivity 0.90, specificity 0.75). Whilst a
Bayesian approach, particularly one based in latent class analysis, to calculating discrimination statistics is appropriate where one is not prepared to presume a gold standard exists, this particular study has limited generalisability secondary to the small sample and single-clinician design, making any inter-rater/agreement comparisons practically impossible.

**B.2.2.7 Methods for identifying risk sub-groups**

The evidence from clinical trials suggests that benefit from DMD therapy may be maximised in patients at higher risk of conversion. Several studies have attempted to formally identify risk sub-groups of CIS patients. A 2009 Cochrane review on the use of DMD for delaying conversion from CIS to CDMS, recommended analysis of treatment efficacy in different patient subgroups in real world clinical practice may help identify those patients at greater risk for conversion. A post-hoc analyses of CIS patients from the CHAMPS trial (intramuscular IFNβ-1a) attempted to identify sub-groups of CIS patients with differing conversion probabilities within the original randomisation based on presenting syndrome and functional system scores at CIS. Following classification into mono-focal and multi-focal CIS presenters, these CHAMPS sub-groups were then re-analysed on the primary end-point of the original trial (conversion to CDMS on Poser Criteria). A greater reduction in the rate of conversion was observed in the mono-focal subgroup (uHR 0.45; 95% 0.27, 0.74) compared with the 36% reduction in the multi-focal presenters (uHR 0.64; 95% CI 0.32, 1.28). However whilst this approach is useful for exploring treatment effect sub-groups, this was a naïve, indirect comparison and entirely unadjusted for other confounding factors. On a more robust, direct multivariate approach, presentation focality was not associated with conversion (p=0.658).

A 2011 French single centre cohort study of both prospectively and retrospectively enrolled patients used an adjusted Cox proportional hazards model to first identify significant correlates of conversion (younger age aHR: 1.44; 95% CI 1.02, 2.01 / spatial dissemination on MRI aHR: 2.07; 95% CI 1.47, 2.91 & >4 WBC/mm3 in baseline CSF aHR: 1.44, 95% CI 1.03, 2.02). The point estimate hazard
ratios from this multivariate model were then used to derive a prognostic score capturing the probability of remaining conversion-free at a particular time point \( t \). The distribution of this prognostic score was then used to divide CIS patients into conversion risk groups (high, intermediate, low). Appropriate analysis to verify the proportional hazards assumption and formal testing for interaction effects within the multivariate model strengthened the methodological validity of the analysis, at least internally.

A 2006 Spanish observational study of serial MRIs from 156 consecutively enrolled CIS patients reported over 5 years of follow-up observed an increasing number of Barkhof criteria at CIS was correlate with an increased hazard of conversion on multivariate Cox (aHR 6.1; 95% CI 2.2, 16.6 for 1-2 criteria, aHR 17.0; 95% CI 6.7, 43 for 3-4 criteria). In addition to predicting differential rates of conversion, categorising CIS patients by the number of Barkhof criteria reported on the baseline scan were then used to aggregate patients into low (zero Barkhof), medium (1-2) and high (3-4) conversion risk groups. Whilst differences were observed in time to conversion between risk groups, the relatively large confidence intervals around the point estimates HRs again suggest larger studies would be required to better localise the independent risk quantum associated with each risk group.

**B.2.3 Conclusion**

What is clear on review of the precedent literature is the sheer number of potential predictors of conversion to CDMS. Under-powering, limited between-study exchangeability and poor isolation of effects secondary to uncontrolled confounding make ranking such predictors in order of the magnitude of their predictive effect or importance difficult. Multivariate modelling was the most frequently reported method for identifying predictors of conversion adjusted for the influence of other known of suspected confounders. Whilst there is support for the theory of differential risk across particular patient sub-groups, prospectively observed, seen-from-onset cohorts have to date
been of insufficient size to confirm these relationships and translate the effect sizes observed in multivariate models into clinically meaningful tools for personalised risk assessment.

### B.2.4 References


B.3 Hypothesis 3: Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict early discontinuation

B.3.1 Search Strategy

References for this review were identified using a search of the PubMed and Scopus electronic databases of peer-reviewed literature using a pre-specified review protocol. The search included the terms “multiple sclerosis” in combination with “treatment discontinuation” or “treatment persistence” within either the article title or abstract published between 2000 and 2013. The references of evaluated articles were then manually screened for additional publications that satisfied the search criteria. The Cochrane review library was then reviewed to identify relevant systematic reviews and meta-analyses not identified in the initial PubMed/Scopus search. The search included references in both English and languages other than English. Case reports were not reviewed. After removing duplicate records, the full reports of each reference satisfying the search criteria were reviewed for relevance and quality to arrive at the final pool of references.

B.3.2 Results

The initial screening consisting of the electronic database search and manual screen of references identified 39 references. A total of 14 references (35.9% of the initial screen) were rejected, 13 on the basis of insufficient relevance and/or quality and 1 duplicate. This left 25 unique references satisfying the inclusion criteria. The PRISMA flow diagram in Appendix I.3 summarises the screen and selection process.

Treatment discontinuation is common in MS.1 Available disease-modifying drugs (DMD) mitigate disease but are only partially effective and on-treatment breakthrough disease is a common occurrence.2,3 Treatment discontinuation is associated with poorer health outcomes,4,5 including significantly higher EDSS and significantly lower proportion of patients remaining relapse and
progression-free compared to patients who remained on DMD therapy. Similar to the literature reviewed around predictors of conversion, the reviewed references report on a wide range of potential correlates of discontinuation including on-treatment events such as relapse and disability progression, lack of efficacy, tolerability issues including side effects and adverse events and demographic and support metrics such as education level and lack of clinician support.

B.3.2.1 Frequentist predictor analyses

Of the 25 references satisfying the review inclusion criteria, the majority explored whether different DMD products, and indeed different doses of the same product, were associated with different rates of discontinuation. There is little supporting data from clinical trials or associated extension studies and/or post-hoc analyses exploring predictors of treatment discontinuation. This is in part secondary to the strict pre-specified treatment stopping rules a priori specified in such trial protocol. As such, whilst such studies clearly have quality advantages with regards to confounder balance and bias control, they are in part somewhat immune to the real-world interaction of factors that drive or promote discontinuation in clinical practice, particular over the long-term.

Given this, it is thus not surprising that some of the best available analyses of treatment persistence are sourced from large real-world observational cohorts sourced from clinical practice. The MSBase Registry itself has previously been analysed for persistence outcomes. A 2013 study of 4054 subcutaneous IFNb treatment discontinuations from 3059 MS patients sourced from the registry reported a higher annual probability of discontinuation in the lower dosage group relative to a propensity-matched higher dose arm (25% vs 20%), following two years of therapy. When translated into a multivariate HR, the lower dose group discontinued at 1.4 times the rate relative to the matched higher-dose group (aHR 1.7; 95% not reported, p<0.0001), adjusting further for time on treatment, age, sex and country. Whilst the study was appropriately powered, a test for underlying hazard proportionality was not reported. A large, prospective sub-study from the MSBase registry of
1113 RRMS patients tracked for a median 2.5 years was the one of the first studies to compare treatment persistence on platform injectables (IFNβ and GLA) relative to natalizumab.\textsuperscript{17} Natalizumab was associated with a 74\% reduction in the rate of discontinuation when compared against intramuscular IFNβ-1a (aHR 0.26; 95\% CI 0.15, 0.45) adjusting for age, sex, disease duration and EDSS. Hazard proportionality was tested and reported (p=0.227). In a separate analysis, older age and longer disease duration at treatment start were both associated with a reduction in the discontinuation rate, however only age remained predictive when adjusted on multivariate modelling (aHR 0.79; 95\% CI 0.71, 0.87). Reason for discontinuation data however was available for just 52\% of the discontinuations analysed, limiting further disaggregation of the Cox models by “reason for discontinuation” group.

Apart from registry database sources, persistence analyses have largely been limited to smaller, single-centre studies. Interestingly, such single-country reports have typically reported a large degree of variance in discontinuation rates which appears to vary as a function of the country, suggesting prescriber country itself may influence persistence beyond other clinical and tolerability factors. As an illustration, a small Serbian study of 290 RRMS patients followed up over 6-years (mean 3.5) reported a far smaller proportion of discontinuation in a study with similar mean follow up as the Jokubaitis analysis (18\% vs 40.3\%).\textsuperscript{17,18} This in part may be secondary to regional differences in attitudes and the therapeutic relationship between clinician and patient. There was significant heterogeneity in treatment persistence by country, with Australia, Canada and the Netherlands all reporting significantly higher hazards of treatment discontinuation when compared with Spain. Consistent with the higher discontinuation rates observed in Canada, a 2011 retrospective cohort study of 682 patients commenced on platform injectables observed high discontinuation rates which were comparable across the 4 products, ranging from 48.8\% for subcutaneous IFNβ-1a up to 54.7\% in IFNβ-1b.\textsuperscript{19} There were no differences between the DMD product groups with regards to either age, residence in long term care or income quintile, however there was
a significantly higher (p=0.04) proportion of females in the glatiramer acetate (GA) discontinuation group relative to IFNβ-1b.

A retrospective German 24-month open-label study of 308 RRMS patients treated with either IFNβ or GA observed a significantly lower 24 month discontinuation rate relative to IFNβ (8.9% vs 32.9%, p<0.001). Whilst this observed difference was large, discontinuation proportions were analysed using a simple chi-square comparison and not adjusted for potential confounders. Furthermore missing data was managed through last observation carried forward imputation which presumes that missing data in thus study was missing either at random (MAR) or missing completely at random (MCAR), which is unlikely in a real world, retrospectively analysed cohort. Outside Western settings and in contrast to the relatively high rates of discontinuation observed in settings such as Northern Europe, Canada and Australia, a 2008 Brazilian cohort study of 1131 MS patients registered in a local Assistance Program for Copaxone (PACO) reported just 10% of glatiramer treatments being discontinued in the 4-year observation period. The most frequently (47%) reported reason for discontinuation was “medical reasons”, although this was not disaggregated further. Whilst this discontinuation rate is indeed far smaller than those reported in comparable studies from a similar time period, this cohort studied were all enrolled in a support program explicitly designed to assist and provide support and guidance for glatiramer users. It is thus not generalizable to broader, non-supported populations of users. Furthermore the analysis, like much of the literature sourced from single centre settings, was limited to simple tests of association with no adjustment made for confounding.

Comparatively few studies extended discontinuation predictor analysis to consider the consequences of ceasing therapy. A 2003 observational cohort study of 1481 patients from 65 clinical centres in northern Italy treated with either intramuscular IFNβ-1a-IM or IFNβ-1b analysed for a three year period between 1996 and 1999 reported worse outcomes in patients who
discontinued treatment in terms of both relapse frequency relative to patients who remained on treatment (ARR 1.67 vs 1.58 for IFNβ-1b, 1.61 vs 1.42 in intramuscular IFNβ-1a) and disability as measured by EDSS (2.5 vs 2.3 IFNβ-1b; 2.4 vs 2.1 in IM IFNβ-1a). Consistent with similar studies featuring inter-product comparisons, basic tests of association (chi-square, t-test) were employed for the analysis. Therefore, no adjustment for systematic differences between stoppers and patients persisting on therapy was performed. This limits the capacity of the study to attribute any difference in disease activity to the actual discontinuation event versus the effects of residual confounding.

Relative to the routine collection of clinical and examination factors such as demography and EDSS, patient and/or clinician reported reasons for discontinuation are less frequently documented and are thus less well studied. Of the small number of reports explicitly documenting reasons for ceasing DMD therapy, a retrospective Irish single-centre chart review of 394 IFNβ treated patients over eight years reported a 5-year discontinuation rate of 28%, consistent with rates observed in southern European settings, but less than the rates observed in Australia, Canada and the Netherlands. Patients ceasing IFNβ therapy secondary to adverse events did so significantly (p=0.0004) sooner in the treatment course (median 13 months; 95% CI 17, 29) compared with discontinuations secondary to treatment failure (median 35 months; 95% CI 32, 45). The use of a simple Wilcoxon rank-sum test, however, meant that such a comparison was entirely unadjusted for systematic differences between the comparison groups. The study, conducted in a single centre and with patients managed by a single clinician, also had limited generalisability. A 2003 Canadian retrospective chart review analysed reason for discontinuation in 844 patients treated with IFNβ. Consistent with the Irish study, the most commonly cited reason for discontinuation was lack of efficacy, accounting for 30% of the 281 discontinuations. Tremlett et al further studied demographic and disease correlates of discontinuation, reporting a significantly higher proportion of females represented in the six-month discontinuations (88% vs 74%, p(chi-square)=0.005) and a higher median EDSS (3.5 vs 3.0, p(rank-sum)=0.008). Consistent with much of the literature around treatment discontinuation from the
early 2000s, no adjustment was made for confounding and no study of the temporal or rate dimension was explored via use of an appropriate time-to-event models.

Whilst much of the literature around treatment discontinuation centres on patients with confirmed MS, recent studies have shifted the focus to isolating independent predictors of treatment discontinuation in CIS patients, i.e. prior to conversion. Early treatment intervention is now considered to be particularly important in high risk CIS patients for delaying or preventing conversion to CDMS, as previously discussed in section 3.1.3.2 above. An exploratory observational cohort study of 1247 DMD treatment episodes from 2314 CIS patients and 44 centres from the prospective, “seen-from-onset” MSBASIS study were studied for independent predictors of discontinuation. Patients were followed for a median 2.7 years and 40.3% of all treatment initiations discontinued within the observation period. Using multivariate Cox PH regression, Meyniel et al reported EDSS change, female sex, location and treatment with IFNβ-1a (either IM or SC preparations) relative to GA as all associated with an increased rate of discontinuation. Hazard proportionality was tested and reported (p=0.07).

B.3.2.2 Non-frequentist methods

Distinct from the analytical approach employed by the majority of the studies reviewed here (basic tests of association, frequentist regression modelling), a 2009 English/Canadian collaboration applied a cost-utility economic evaluation approach in an attempt to identify the magnitude of disability at discontinuation that improves overall DMD cost-effectiveness. Using Medicare data to simulate cost and health outcomes for seven active treatment options, which included a range of discontinuation triggers, the study estimated that if disability progression was, theoretically, the only determinant of discontinuation then a treatment strategy that ceased therapy upon progression to EDSS of 7 would maximise overall cost-effectiveness, in terms of a gains to cost ratio. Of course discontinuation is rarely this simple and one-dimensional.
Beyond traditional metrics and factors collected and measured in routine clinical practice, one 2003 study focused on behavioural and/or attitudinal correlates of treatment discontinuation. This study of 946 MS patients treated with IM-IFNβ-1a employed a questionnaire and interview approach to populate a transtheoretical behavioural model of change. Then model was designed to best differentiate patients who discontinued therapy from those who persisted, based on both attitudes towards the drug in addition to traditional predictors of discontinuation such as disability level. Berger et al’s model found that discrimination was maximised in a model that incorporated a patient’s attitude or perception of the advantages and disadvantages of this DMD product, education level and EDSS – correctly identifying 82% of discontinuations and 81% of stayers. Whilst such a behavioural trans-theoretical approach has not been externally validated in CIS patients, it does underscore the broader contention that treatment discontinuation, like the relapse onset timing and conversion to CDMS end-points examined in hypotheses 1 and 2, are complex and multifactorial.

B.3.3 Conclusion

Previous studies have identified a diverse range of patient, disease and treatment characteristics that correlate with DMD treatment discontinuation including sex, country, on-treatment relapse activity and EDSS change, disease duration and education level. Further, differences in persistence have been previously observed between different immunomodulatory therapies. However the performance of these characteristics as reliable and replicable predictors of treatment persistence have not before been examined in a real-world, prospectively assessed and seen-from-onset cohort of this size and global spread previously. From an analytical perspective, the task of isolating the independent influence of any one predictor from a mess of other influences, confounders and design or methodological biases requires the use of large datasets and appropriate statistical methodology to minimise bias and confounding in estimating true effect.
The paucity of appropriately-powered, prospectively observed, real-world data examining the characteristics, either demographic or disease-activity based, of patients who cease treatment makes identification and, importantly, anticipation of patient subsets likely to demonstrate sub-optimal response to currently available platform therapies difficult. This is particularly important early on in a patient’s treatment course where discontinuation rates have previously been observed to be as high as 20% within the first 6 months of DMD therapy and up to 70% within the first year of treatment in some settings.  

B.3.4 References


Timothy Denis Spelman (58172)


B.4 Hypothesis 4: Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment scenarios and products

B.4.1 Search Strategy

References for this review were identified using a search of the PubMed and Scopus electronic databases of peer-reviewed literature using a pre-specified review protocol. The search included the terms “multiple sclerosis” in combination with “propensity score” within either the article title or abstract published between 1995 and 2014 inclusive. The references of evaluated articles were then manually screened for additional publications that satisfied the search criteria. The Cochrane review library was then reviewed to identify relevant systematic reviews and meta-analyses not identified in the initial PubMed/Scopus search. The search included references in both English and languages other than English. Case reports were not reviewed. After removing duplicate records, the full reports of each reference satisfying the search criteria were reviewed for relevance and quality to arrive at the final set of eligible reference.

B.4.2 Results

The initial screening consisting of the electronic database search and manual screen of references identified 78 unique records. A total of 33 references were rejected, 25 on the basis of insufficient relevance and/or quality and 8 as duplicates. A total of 45 references were included in the final review. The PRISMA flow diagram in Appendix I.4 summarises the screen and selection process.

Disease-modifying drugs are usually approved on the basis of placebo-controlled randomized studies. Therefore, DMD switch decisions comparing two active drugs and many head-to-head first-line product comparisons are not supported by level 1 evidence studies. Prospective MS outcome datasets acquired in the observational setting through registries or phase IV studies represent another opportunity to generate efficacy comparisons between DMD products not available from
the RCT evidence base. Observational studies are, by definition, non-randomized, with treatments chosen as a result of patient and physician preference and the applicable regulatory environment. Therefore, comparison groups may vary widely from each other with regard to their pre-treatment characteristics, leading to selection bias and confounding. Such imbalance can fatally confound any demonstrated difference in efficacy by treatment arm, limiting the ability of the researcher to be able to attribute such difference to the treatment of interest and potentially compromising subsequent treatment decisions in the clinical setting. Propensity score (PS) matching is a statistical technique for correcting for covariate imbalance in non-randomly selected cohorts. Propensity based methodologies, with their ability to balance the distribution of known or suspected confounders of treatment efficacy across study arms, can reduce selection bias and have been shown to closely approximate estimates of treatment effect derived from randomized trials whilst preserving the real-world characteristics of observational registry data.

The appropriate use of PS methods can allow investigators to robustly approximate the common time-to-event metrics of treatment effect reported by controlled clinical trials. Estimation of treatment effects via PS methodology may in addition be hypothesis-generating, providing guidance for later RCTs.

**B.4.2.1 Propensity score matching methodology & performance**

The contemporary theoretical literature supporting application of propensity score methods in head-to-head medical treatment or intervention comparisons owes a significant debt to the series of simulation studies and reports by Austin et al. A central theme in Austin’s various simulation studies is a comparative analysis of the relative performance of a variety of PS-based techniques including propensity score matching (PSM), stratification, weighting and covariate adjustment in estimating a true treatment effect when sampling from an entirely simulated, and thus known, population. Austin’s 2013 simulation study employed Monte Carlo simulation to compare PS matching, stratification, inverse probability of treatment weighting on the PS and covariate adjustment on the estimation of a time-to-event marginal hazard ratios in a head-to-head,
treatment vs comparator design. Of these four PS based methods, the weighting method minimised the model residuals (observed minus estimated) relative to the other three methods trialled whilst both weighting and matching on the PS resulted in unbiased estimation of the true treatment effect. A prior 2008 study also using Monte Carlo simulation to compare the performance of these different PS methods for estimating, in this case, relative risks, similarly found PS matching returned a less biased estimated of the true treatment effect relative to stratification on quintiles of the PS. Matching on the PS also appears to out-perform stratification or weighting to balance treatment and comparator study arms. Another 2008 simulation study similarly observed that systematic differences between unmatched treatment arms in baseline confounders were minimised under PS matching, relative to any of weighting, stratification or covariate adjustment on the PS. This analysis compared the performance of matching treatment arms via propensity score adjustment relative to stratification on quintiles of the propensity score and standard covariate adjustment in the estimation of relative risks using a similar Monte Carlo simulation. Across all simulations, propensity score matching again out-performed either stratification or multivariate covariate adjustment in terms of minimising bias (the difference between the true and estimated effect).

This is consistent with an earlier simulation study by Austin comparing PS matching, stratification and covariate score adjustment in estimating odds ratios (rather than relative risks or hazard ratios) where the matching approach again produced an unbiased estimation of the true treatment effect whilst stratification resulted in a minor bias. Quintile stratification in this simulation, however, returned more precise estimates relative to either matching or regression adjustment. Importantly, both propensity score matching and stratification clearly outperformed traditional multivariate regression covariate adjustment in terms of minimising bias and maximising precision. Indeed across this series of simulation studies Austin concluded that propensity score adjustment, either via matching or stratification, returned relatively unbiased estimation of odds ratios, relative risks or
hazard ratios. This trend from the theoretical literature generally supporting matching over covariate adjustment was corroborated by similar studies in 2007 and 2009 which likewise observed baseline imbalance between study arms was minimised under a propensity matching procedure relative to either score adjustment in addition to, in this particular study, stratification.14,15

Outside Austin’s work, Cepeda et al’s 2003 Monte Carlo-based simulation study comparing the bias, precision and robustness of quintile stratification based on propensity score against a straight (non-PS adjusted) logistic regression observed superior control of confounder imbalance and a reduction in bias under PS stratification, particularly where the end-point of interest was rare.16 Lee et al favoured classification analysis over simulation for performance testing, employing a machine learning extension of propensity score derivation via classification and regression tree (CART) analysis.17 Whilst the methodology was quite different from that employed by Austin and others the results were consistent, with PS matching again observed to minimise bias relative to either covariate adjustment or weighting on the PS.

Beyond direct comparisons between alternate PS-based techniques, matching on propensity score has been further observed to out-perform conventional non-PS based matching procedures in terms of both balancing confounders between treatment arms,18 to the extent where the quality of PS matching has been observed approximate the degree of balance observed under random assignment within clinical trials.19 Of the reviewed literature only a single study found no advantage favouring propensity score adjustment over standard multivariate adjustment. A systematic review of 69 medical studies between 1998 and 2003 that employed propensity scores adjustment in some form observed, in general, that treatment effect estimates derived from analyses employing PS methods were not markedly different from the equivalent effect sizes estimated from conventional multivariate models employing standard covariate adjustment.20 This review, however, was limited by significant heterogeneity between included studies and incomplete details around the exact
propensity score based method used. Furthermore, no formal assessment of bias, such as those derived in the various simulation studies, was presented.

The capacity to correct treatment arms for imbalance, at least with regards to those variables used in the derivation of the propensity score, makes propensity score adjustment an attractive option for conducting head-to-head treatment comparisons in real-world observational cohorts. Whilst the application of PS adjustment has become increasingly popular in the medical literature over the past 15-20 years, it previously found broad use within the health economic evaluation literature, primarily as a tool for minimising selection bias when inputs used by the economic modelling were sourced from non-randomised pharmaco-epidemiological studies. They have also been previously used for managing bias and confounding in the analysis of large administrative health care and claims datasets when studying treatment outcomes and events in a post-marketing setting such as drug toxicity, mortality, or adverse events. A 2006 review of the use of propensity scores in pharmaco-epidemiological studies identified several advantages of propensity score adjustment over standard multivariate regression including superior management of indication confounding, particularly when analysing real-world data sourced from clinical practice settings, including registry data.

The presence of significant imbalance by treatment arm in those prognostic correlates of treatment outcome, which also tend to predict which DMD product a patient is prescribed, is a considerable challenge in head-to-head efficacy comparisons analysed from observational cohorts. In conclusion, standard multivariate covariate adjustment does not adjust for such bias whilst a propensity match approach can at least balance treatment arms with respect to the drivers of such indication bias.

Furthermore, close study of unmatched individuals in the “tails” of the distribution of the propensity scores in the context of real-world treatment outcome comparisons can be instructive regarding
what types of patients approach absolute indication or contra-indications to treatment in the real world. Further, by distilling a large number of effect confounders into a single propensity score metric a propensity-adjusted analysis, either via matching, stratification, weighting or covariate adjustment, can markedly improve both the precision and accuracy of treatment effect estimates, particularly where the study end-point is infrequent.\cite{16} By comparison, a traditional multivariate regression based on maximum likelihood estimation performs optimally only under a sufficient number of outcomes (e.g. minimum 8-10 outcomes per explanatory variable included in the model).\cite{26} Thus propensity score adjustment can confer greater explanatory power and flexibility to the analysis.

B.4.2.2 Propensity score adjustment in multiple sclerosis

A total of 15 of the included references studied the application of propensity score methodology specifically to MS data. A number of recent MS studies have employed PS-based methods for refining estimates of treatment effect in absence of equivalent head-to-head comparisons from clinical trials. Trojano et al, pioneers in the application of PS adjustment methods to observational MS data, used a Cox proportional hazards regression adjusted for PS across 2090 MS patients from the Italian MS Database Network to observe a significant reduction in both the rate of disability progression (aHR 0.23, 95% CI 0.17, 0.30) and rate of relapse (aHR 0.19, 95% CI 0.14, 0.17) in patients on IFNβ therapy for a minimum of four years relative to a comparator of shorter (up to 2 years) IFNβ treatment duration.\cite{27} The same study group subsequently employed a variant of this approach, using inverse weighting on the PS rather than fixed effect covariate adjustment in the Cox model to evaluate the impact of IFNβ treatment on long term disability changes in an observational cohort of 1103 treated RRMS patients against an untreated comparator arm (n=401).\cite{28} After adjustment for PS, the IFNβ arm demonstrated a reduction in reaching an EDSS 4.0 milestone from first recorded visit (HR 0.69; 95% CI 0.52, 0.93), EDSS 6.0 (HR 0.60; 95% CI 0.38-0.96) and time to secondary progressive MS (SPMS) (HR 0.38, 95% CI 0.24, 0.58) all relative to the PS-weighted
untreated comparator arm. Subjects in this analysis were followed for up to 7 years, well beyond the typical 12 to 36 month duration of MS clinical trials and associated extension studies. This highlights an advantage of studying observational cohorts – the ability to study long-term outcomes, particularly relevant in a chronic disease such as MS, where patients are typically on treatment for years if not decades. Similarly, a 2009 Italian analysis of 2570 RRMS patients on IFNβ therapy followed for up to seven years used PS quintiles to adjust the time-to-event Cox PH model. This study found early IFNβ treatment significantly reduced the rate of reaching an EDSS of 4.0 (PS adjusted HR 0.56; 95% CI: 0.36, 0.90) and a single point progression in EDSS (aHR 0.63; 95% CI 0.48, 0.85).

An earlier 2004 retrospective study of 695 RRMS patients sourced from 22 Italian centres employed PS covariate adjustment of a Poisson regression model to test whether the efficacy advantages of platform injectable IFNβ therapy observed in randomised trials were replicable in contemporaneous real-world, clinical practice settings. Deriving PS conditional upon baseline clinical and demographic factors a priori identified as likely to correlate with the clinician decision to initiate IFNβ treatment, Russo et al compared the risk of disability progression in a treated cohort against an untreated comparator via several methods: 1) unadjusted univariate regression, 2) standard fully adjusted multivariate regression. 3) PS adjusted regression model and 4) full multivariate model further adjusted for PS score. Whilst all models returned a significant reduction in the risk of progression favouring IFNβ treatment in the 2-4 years treated sub-group, the size of the estimated average risk reduction was largely consistent across the four approaches from a 60% reduction under univariate (unadjusted) modelling (RR 0.40; 95% CI 0.24, 0.66) to a 65% reduction using PS methods, suggesting unadjusted analysis may, in this case, underestimate the true treatment advantage of IFNβ, if we accept that PS matching provides for superior control of systematic imbalance between treatment arms.
Propensity score methods have also been used in post-marketing MS studies. A 2008 phase IV study of the effect of neutralizing antibodies on IFNβ efficacy in RRMS patients observed a significant increase in the both the risk and incidence of relapse in the antibody positive group relative to the propensity-matched antibody negative arm. A post-marketing analysis of the same Italian cohort combined PS matching with a tree-growing recursive partitioning technique (a variety of cluster analysis) to explore a possible gender differential in response to IFNβ therapy, observing male sex was associated with a decreased rate of first relapse event relative to females (HR 0.86; 95% CI 0.76, 0.98) and, conversely, an increased risk of a one point EDSS progression (HR 1.33; 95% CI 1.00, 1.76, p<0.05). Such an approach illustrates the flexibility and adaptability of PS-based techniques, particularly for supplementing other non-PS based analytical approaches.

A number of registry-based studies have used propensity methods to reduce confounding in large observational cohorts. Kalincik et al used PSM to correct for differences in the distribution of prognostic correlates of DMD treatment persistence, relapse and disability progression outcomes between two dose groups of IFNb-1a-SC sourced from MSBase. Age, disease duration, annualised relapse rate (ARR), EDSS, disease course, MRI activity and a clinic identifier were used to derive the propensity score which was then applied to match 610 patients on lower dose SC-IFNβ-1a (22 micrograms) on a 1:1 basis with 610 comparable MS patients in the larger dose (44 micrograms) group. Whilst the lower dose group were more likely to discontinue relative to the matched high dose group (25% vs 20% annual discontinuation probability), there was no difference in either relapse or disability progression rates by dose group. Importantly, the application of PSM meant these observations were not confounded by systematic differences in those baseline factors used to derive the PS score. Pasternak et al similarly employed PSM to interrogate and link the Danish Prescription Drug Registry with the Danish National Registry to identify 36659 amiloride users and match these on a 1:4 basis with a pool of 177031 thiazide diuretics users to study incident MS. Age, sex, calendar year, socio-economic class, degree of urbanisation, co-medication, any cause
hospitalisation and the Charlson comorbidity index were all incorporated into a logistic regression model to calculate a patient-level propensity score which was then used to balance the comparison groups. Through balancing the matched sample for these factors, amiloride use was not observed to correlate with a change in incident MS risk, independent of any influence from the factors used in deriving the propensity score.

Similar methods have been used to analyse smaller, generally single-centre datasets with mixed results. A 2013 Canadian cohort study of n=432 mothers with MS used PSM to compare duration of birth hospitalisation in MS mothers with a frequency matched sample from the general perinatal population sourced from the British Columbia Perinatal Database Registry. Using PSM to control for systematic differences in confounders of length of stay, no difference in the duration of birth hospitalisation was observed in MS mothers relative to the general perinatal population. A 2012 retrospective cohort study of 2656 MS patients, also sourced from British Columbia Multiple Sclerosis database during a 1985-2004 study period, similarly employed PSM to minimise baseline confounding when comparing time to confirmed disability progression in patients treated with IFNβ relative to both a contemporaneous untreated cohort and a historical untreated comparator. Using age, sex, disease duration, baseline EDSS, ARR, SES and Charlson comorbidity index to derive the PS, Shirani et al observed no significant difference in progression to an EDSS milestone of 6 in the matched IFNβ group relative to either the contemporary untreated controls (HR 1.34; 0.93, 1.92) or the historical controls (HR 0.84; 95% CI 0.63, 1.11). These results were consistent with the primary analysis using full multivariate covariate adjustment in place of PSM. A 2012 analysis of patient reported outcome and clinician assessment metrics from 1082 MS patients from a single Cleveland neurology clinic used propensity scores to stratify the sample into quintiles of the PS for early treatment. By calculating the theoretical probability of initiating early treatment conditional on age at symptom onset, sex, ethnicity, follow-up time and relapse activity, the investigators estimated patients in the early treatment arm were associated with significantly higher EuroQOL-5D, Patient
Further illustrating its analytical flexibility, propensity score methods have also been combined with clustering methodology to help localise treatment effects in MS studies. Goodin et al approached the common problem of end-point dependent exposure variation in analysis of long-term MS treatment outcomes in non-randomised cohorts by combining weighted stratification on the PS with a recursive-partitioning algorithm. Such exposure variation is a form of informative censoring where decisions to initiate, continue or cease DMD treatment may be influenced by systematic differences in patient characteristics between treatment and comparator arms, thereby introducing a treatment-selection bias which can disguise or obscure genuine treatment effects on long term MS outcomes. Sourcing data from 260 patients enrolled in the US/Canadian Betaseron 16-Year Long Term Follow-Up study, the investigators doubly weighted patients, first by cumulative treatment exposure duration and then for various confounders of long term response to treatment. Recursive-partitioning was then used to identify the optimal weighting scheme - i.e. the scheme most strongly associated with the primary end-point (time to reaching EDSS>=6 milestone). This permitted the identification of weighted high and low exposure groups which were then stratified based on the PS to further adjust for other baseline imbalance. Using this methodology, sustained treatment with IFNβ-1b was associated with a significant reduction in the rate in any of the outcomes studied (EDSS=6, EDSS=6 or SPMS, SPMS, Wheelchair, Any) HRs ranging from 0.30-0.42 (all p<0.0001).

Combining propensity score derivation with regression-based calibration to correct for possible measurement error often associated with observational data has been flagged as an alternate approach to quantifying the magnitude of any hidden bias secondary to unmeasured confounders. In a 2007 simulation study, Strumer et al estimated the performance of a theoretical propensity model in estimating a treatment effect in both the presence and absence of the key “no unmeasured
confounders” assumption. Where the assumption was satisfied, that is where a gold-standard propensity score contained no additional information on the outcome compared to the actual calculated PS, then a PS calibration approach lead to a reduction in bias of between 32% and 100%. However, when this surrogacy was violated, then a calibration approach was found to potentially increase bias. This underscores the importance of critically appraising whether the surrogacy assumption is sufficiently satisfied when preferencing a calibration approach over straight PS matching, stratification, weighting or covariate adjustment.

PS methods have also been employed to test observational cohorts for the presence of significant imbalance. A 2011 prospective analysis of 15 years of data from the open-label US Glatimer Acetate Trial used propensity score analysis not to adjust the EDSS change end-point comparisons, but rather as a pre-analysis screen for potential confounder imbalance between the different patient subgroups. No difference in the distribution of propensity scores was observed between the “ongoing” sub-group (both patients originally randomised to GA in the double-blind phase of the trial and persisting following the switch to open-label phase and those originally randomised to placebo but who switched to GA once the open-label phase started), the “withdrawn” sub-group (patients who withdrew from the study at some point following at least 1 dose of GA) and a third, composite modified intention-to-treat (mITT) population combining both ongoing and withdrawn groups. By statistically establishing balance at by PS analysis (PS distribution: Ongoing = 0.47 +/- 0.12, Withdrawn = 0.40 +/- 0.12 & mITT = 0.43 +/- 0.12, all p>0.05), the investigators were then not required to PS adjust the subsequent comparative analysis. However, this PS-based assessment of balance was made at baseline only and the influence of post-baseline events on sample balance, such as relapse, was not assessed.

Propensity score covariate adjustment has further been employed to study pregnancy and foetal outcomes following exposure to IFNβ in MS patients. A 2010 study of 396 pregnancies from 388
women with MS from 21 Italian centres derived a propensity score conditional upon a large series of perinatal/obstetric history factors (age at conception, previous pregnancies, delivery mode, gender of baby, gestational age and alcohol and drug exposure during pregnancy), MS disease characteristics (EDSS, MS disease duration) and socio-economic metrics (educational level). Quadratic terms were also included in the model deriving the propensity score to test for potential non-linear associations between such factors and IFNβ exposure. Two separate applications of the propensity score were then applied. First, the distribution of propensity scores in the IFNβ exposed treatment and non-exposed comparators were studied and subjects with non-overlapping PS excluded from the analysis dataset. Secondly, the models used to compare outcomes between groups were then further adjusted on PS quintiles, to estimate IFNβ exposure within 4 weeks of conception was associated with both a significant decrease in baby weight (linear model $\beta= -1113.8$, $p<0.0001$) and length at birth ($\beta= -1.102$, $p<0.0001$) relative to controls. Whilst there is increasing evidence supporting PSM over covariate adjustment and even stratification, this two-step amalgam of excluding non-overlapping scores coupled to quintile adjustment was not compared to a formal PSM procedure, arguably a superior method.

PSM has also been used to study pharmacy claims data in a program evaluation setting. Stockl et al analysed 8-months of MS injectable medication dispensation data to assess both treatment persistence and adherence in a cohort of MS patients enrolled in a MS disease therapy management (DTM) program aimed at improving patient self-management in the United States. PSM was used to match patients from the DTM program with two comparable non-program control arms sourced, alternately from retail and speciality pharmacy database sources. The PS was derived using a logistic regression where theoretical likelihood of enrolment in the DTM program formed the dependent outcome variable specified on a function of age, sex, DMD product, chronic disease score and health plan. Analysis of the PS matched sample demonstrated DTM was associated with an increase in adherence relative to the non-program retail pharmacy cohort (0.92 vs 0.90, $p<0.001$) and a greater
persistence (220 days) relative to either retail (188 days, p=0.002) or speciality pharmacy (177 days, p<0.01). However, the propensity score used in this analyses did not include other key potential confounders of either adherence or persistence including baseline EDSS, disease duration, previous DMD exposure and pre-program relapse activity. Thus the observed differences may remain subject to confounding secondary to systematic differences in these factors between groups. No sensitivity analysis was conducted to estimate the effect of unmeasured confounding.

PS methods have further been employed to control for confounding in comparisons of MS treatment switch or escalation. A 2012 post-marketing, prospective study of 267 RRMS patients from two Italian centres used Cox proportional hazards modelling inversely weighted on the propensity score to compare relapse, progression and MRI activity end-points in an observational cohort of patients switching to an alternate IFNβ/GA after failing first-line IFNβ/GA relative to a cohort escalating from IFNβ/GA to natalizumab.\(^\text{42}\) On application of PS weighing, the adjusted escalation group was associated with a decreased rate of 24-month relapse, progression and MRI activity (adjusted HRs ranging from 0.38 to 0.56 where switching was the reference group). Whilst C-statistics were used to test the validity of the model used to derive the PS, an alternate matching approach was not used despite the weight of evidence from the key simulation studies supporting PS matching over weighting. However, unlike many of the studies reviewed herein relying on some variety of PS adjustment, the influence of unmeasured confounding was formally tested via a Greenland sensitivity analysis. In this case, a theoretical unobserved confounder associated with a minimum relative risk of 1.5 on a background of a minimum prevalence imbalance between switch and escalation groups of 60% did not significantly alter the observed treatment effect in favour of selection bias. Similarly a separate Greenland sensitivity presuming a minimum RR estimate of 4.0 and 25% imbalance found no evidence of significant bias secondary to unmeasured confounding. Arguably, these assumptions only exclude unknown confounders with very large effect sizes.
B.4.2.3 – Reporting and evaluation

What is broadly consistent across the reviewed literature is the generally inadequate reporting of the relevant checks and balances of both PS method assumptions and performance, including the frequent absence of balance and bias statistics necessary for objective assessment of how well the PS-based approach worked. Almost all of the studies reviewed herein did not reference the key “no unmeasured confounders” assumption. Only two studies conducted a formal test or sensitivity analysis to estimate influence of unmeasured confounding.\(^7\)\(^,\)\(^42\) Indeed, the lack of a minimum standard in reporting PS-adjusted analyses appears to be a problem not contained to the MS literature. A 2007 systematic review of the use of PS matching in studies of cardiovascular surgery outcomes identified 60 studies between 2004 and 2006.\(^21\) Of these, only 31 (51.7%) provided sufficient information regarding how the score was derived and patients subsequently matched. Not one of the 60 studies used appropriate tests of difference for comparing baseline characteristics by matched treatment arms- favouring unpaired over paired tests, thus ignoring the paired structure of the matched cohort. The authors recommended, as a minimum, presenting a comparison of baseline factors between matched arms and reporting the tests used to assess balance. These findings were consistent with a 2010 systematic review of 47 studies across 2006-2009 in the intensive care and anaesthesiology literature that similarly employed some variety of PS analysis.\(^43\) Of the reporting deficiencies noted, the authors made particular reference to the lack of formal balance statistics and inadequate description of the model structure used to derive the propensity score. Consistent with these observations, a 2010 review of medical studies employing PS methods found that only a single study of the 27 reviewed applied some variety of sensitivity analysis to test the performance of PS adjustment.\(^8\)

Whilst it may be tempting to consider PS adjustment as a panacea for the various biases often plaguing real-world observational data, the subsequent analysis used to analyse PS adjusted data to
estimate treatment effect needs to be appropriate for the matched design of the dataset. Indeed the 2010 Austin review of the cardiovascular surgery literature described in the above paragraph further recommended that PS-based studies should report an appropriate analytical method for studying the end-point across treatment arms which explicitly accounts for the matched design (e.g. a Cox Marginal Model in time-to-event end-points). Indeed, the application of an inappropriate model or method to PS adjusted data is the single most frequent error observed in the reviewed literature. Austin highlighted this problem in another review (2014) which observed that PS methods were, in particular, incorrectly applied to time-to-event outcomes including the use of unpaired tests to assess balance in the matched sample and the estimation of standard HRs rather than the more correct marginal HRs appropriate for a matched design. In addition, recent systematic reviews of estimating statistical significance in PS-adjusted data found that the matched nature of the analysis dataset is routinely ignored when estimating the significance level of the treatment effect, usually resulting from the application of a model that similarly ignores the matched design. This was consistent with Austin’s 2009 simulation study showing the same.

With regards to the importance of reporting the suite of explanatory variables included in the model used to derive the propensity score, Brookhart et al demonstrated, by simulation, that the exact choice of included factors can have important downstream effects on the variance and residual bias in the treatment effect estimation. Specifically, the investigators found that the inclusion of covariates that were related to the treatment exposure but only weakly correlated with the outcome or end-point of interest (e.g. relapse, disability progression) can result in significant increases in the variance around the point estimate of treatment effect. Furthermore such weakly correlated factors not only increased effect size estimate variance and thus, decreased the precision of the point estimate, but also tended to remove only marginal, and frequently negligible bias from the cohort. This supports not only stating the covariates used to derive the propensity score but also providing a brief rationale for their inclusion.
B.4.3 Conclusion

In the absence of randomized clinical trials, propensity-matching techniques can estimate the benefits associated with various treatment decisions in a clinical practice setting. Unlike randomized controlled trials, propensity score based approaches cannot adjust for confounder imbalance in baseline characteristics that have not been recorded. However, propensity matching as a variety of pseudo-randomisation has been demonstrated to both reduce selection bias and closely approximate the risk estimates derived from randomized trials.

B.4.4 References


10. Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine, 33*(7), 1242-1258.


44. Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine, 33*(7), 1242-1258.

C. METHODS

C.1 The timing of relapse onset is seasonal and this relationship is latitude-dependent - methods


Authors:
Tim Spelman¹,², Orla Gray³, Robyn Lucas⁴ and Helmut Butzkueven¹,² on behalf of the MSBase Study Group.

Authors: institution(s)/affiliation(s) for each author:

Tim Spelman

¹ Department of Neurology, Royal Melbourne Hospital, Parkville, Australia

² Department of Medicine (RMH), The University of Melbourne, Parkville, Australia

tim@burnet.edu.au

Orla Gray

³ Department of Neurology, Ulster Hospital,
Timothy Denis Spelman (58172)

Belfast, Northern Ireland

orlagray@hotmail.com

Robyn Lucas

National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia

Robyn.Lucas@anu.edu.au

Helmut Butzkueven

Department of Neurology, Royal Melbourne Hospital, Parkville, Australia

butz@unimelb.edu.au

Corresponding author:

Tim Spelman

Department of Neurology, Royal Melbourne Hospital, Parkville, Australia

Department of Medicine (RMH), The University of Melbourne,
Keywords:
Multiple Sclerosis, relapse, residual plots, trigonometric regression, sine regression, seasonality, latitude

Short abstract:
Combining plot analysis with trigonometric regression is a robust method for exploring complex, cyclical phenomena such as relapse onset timing in multiple sclerosis (MS). This method enabled unbiased characterisation of seasonal trends in relapse onset permitting novel inferences around the influence of seasonal variation, ultraviolet radiation (UVR) and latitude.

Long abstract:
This report describes a novel application of trigonometric regression modelling to 55 years of multiple sclerosis relapse data from 46 clinical centers across 20 countries located in both hemispheres. Central to the success of this method was the strategic use of plot analysis to guide and corroborate the statistical regression modelling. Initial plot analysis was necessary for establishing realistic hypotheses regarding the presence and structural form of seasonal and latitudinal influences on relapse probability and then testing the performance of the resultant models. Trigonometric regression was then necessary to quantify these relationships, adjust for important confounders and provide a measure of certainty as to how plausible these associations were. Synchronization of graphing techniques with regression modelling permitted a systematic refinement of models until best-fit convergence was achieved, enabling novel inferences to be made regarding the independent influence of both season and latitude in predicting relapse onset timing in MS. These methods have the potential for application across other complex disease and
epidemiological phenomena suspected or known to vary systematically with season and/or geographic location.

**Introduction:**

The most common form of multiple sclerosis (MS) is Relapsing Remitting Multiple Sclerosis (RRMS). RRMS is characterized by episodic deteriorations in neurological function, followed by partial or complete recovery. Globally, the incidence and prevalence of MS increase with increasing distance away from the equator in both hemispheres.\(^1\)\(^-\)\(^3\) Whether the frequency of relapse events that occur specifically in RRMS also vary with latitude, and whether there is any underlying seasonal variation in any such association, remains unclear. To date studies exploring seasonality in relapse timing have been limited to single clinical centers, limiting any inferences regarding seasonal trends in relapse timing to solitary geographical locations and thus unable to explore broader latitudinal influences.\(^4\)\(^-\)\(^14\) These studies have been further limited by small sample sizes and sparse relapse data. A 2000 meta-analysis of ten studies from clinical centres in Europe, the United States and Canada, where each study included a minimum of thirty cases reporting the season-of-onset of relapses, described a clear seasonal trend in the timing of relapse onset, with relapses peaking in spring and with a winter trough\(^5\). Similar cyclical annual trends in onset have been observed in subsequent, albeit smaller, studies in both Japan\(^15\) and Spain\(^16\). However, a comparable United States study failed to corroborate this pattern\(^17\). To date, these studies and observations have been limited to the northern hemisphere. The MSBase study group recently analyzed a large global dataset of MS relapses across both northern and southern hemispheres to explore seasonal trends in the timing of relapse onset in addition to latitudinal influences on the relationship between peak relapse probability and seasonal ultraviolet radiation (UVR) trough\(^18\). Central to these methods was the application of trigonometric regression to visualize and evaluate trends in the timing of relapse onset and UVR distributions.
The overall goal of this study was to test the hypothesis that temporal variation in the timing of relapse onset in MS varied predictably with season in both the northern and southern hemispheres and this seasonality was influenced by latitude. The rationale for the use of trigonometric modelling to investigate these questions was its flexibility for characterizing two- or three-dimensional phenomena that are known or suspected to describe discrete, predictable and consistent shapes or patterns, such as the annual cycle of peaks and troughs commonly observed in biological or epidemiological phenomena possessing seasonality.\textsuperscript{19-22} A disadvantage of conventional time-series analyses, including Fourier analysis, is the presumption that time series are often characterized by stochastic processes.\textsuperscript{21,23,24} By contrast, incorporating trigonometric functions into a regression type model has the advantage of both facilitating exploration of regular and systematic structures in periodic data whilst exploiting the regression model structure to explore other correlates or adjust for confounders of seasonality.

Trigonometric regression has previously been used widely in the medical epidemiological literature to explore temporality in topics as diverse infectious disease outbreak detection, the role of circadian rhythms in everything from autonomic nervous system dysfunction to preterm placental abruption through to seasonal correlates of congenital malformations and the timing of presentations of accident and emergency.\textsuperscript{25-32} Such modelling typically demands larger sample sizes than more conventional time-series analyses and as such this is the first time it has been applied to a global dataset of MS relapse onset. Trigonometric regression as described here is suitable tool for investigators exploring any phenomena which is known to or suspected of cycling systematically over time. Not only can such modelling help characterise and visualize these patterns, it further permits the user to explore potential drivers and correlates of these trends.

Regarding the specific example of MS relapse onset presented here, the use of scatter and residual plots to visualize and assess how closely a hypothesized geometric model form fits the data
constitutes the critical step in determining: 1) whether the observed data provide sufficient evidence to support a hypothesis of seasonality or other temporal trends in the timing of relapse onset; and 2) whether the frequency and arrangement of sine and cosine functions which define a particular trigonometric model is adequate to permit use of this model for subsequent inference and prediction. Regression modelling also permits control for important confounders of any observed seasonal or latitudinal effect such as patient-level propensities for relapse, particularly factors which in themselves are time-varying such as the duration of pre-relapse exposure to disease-modifying drug (DMD) treatment. Isolating independent geographic and temporal predictors and correlates of relapse onset timing in MS has the potential to guide biological investigation into the mechanisms of relapse events which in turn may inform the development of future treatment interventions aimed at preventing or delaying disease exacerbation.

The MSBase Registry

MS patients contributing relapse data to this analysis were sourced from the international MSBase registry. Established in 2004, the registry longitudinally collates demographic, disease activity, clinical examination and investigation characteristics and metrics from consenting patients attending MS clinic using an internet-based, physician-owned and operated system www.msbase.org. Member centers follow a common protocol that defines the minimum dataset required to be uploaded at agreed regular intervals to ensure outcome data such as relapse events are consistently and prospectively compiled. The date of relapse onset is included as a mandatory minimum dataset variable. In addition relevant clinical data associated with these relapse events is commonly collected including corticosteroid treatment and functional system affected. The use of the common iMed data entry system further ensures a unified approach across centers to data collection and reporting. This project holds Human Research Ethics Committee approval or exemption at each contributing center. Informed consent according to local laws from all patients included in the analysis is mandatory.
Inclusion criteria
A total of 9811 patients contributing 32,762 relapse events were included in the analysis. Clinical MS centres with a minimum of 20 registered patients consented, uploaded and tracked in the registry as of the 1st December 2013 (date of data compilation) were eligible for inclusion in the analysis. To ensure all relapse events included in the analysis were prospectively observed, only relapse onsets dated subsequent to the first recorded patient disability assessment (using the Kurtzke Expanded Disability Status Score (EDSS)) were included in the analysis. All patients contributing relapse data to the analysis satisfied formal diagnostic criteria for MS.\textsuperscript{34,35}

Outcome measures
This study considered two primary outcomes: 1) whether there was temporal variation in the probability of relapse onset at the level of the geographic location, the hemisphere and/or globally; and 2) whether there was a relationship between latitude and the lag, in months, between the timing of seasonal UVR trough and the subsequent peak relapse probability date. The MSBase Study group hypothesized that as absolute vitamin D levels are lower in regions further away from the equator and location-specific seasonal population level vitamin D nadirs are likely reached earlier following the winter solstice in such distal locations, then the effect of low vitamin D levels on increased MS relapse probability would similarly describe such temporal and latitudinal patterns.

Relapse definition and dates
A relapse was defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 hours, in the absence of concurrent illness or fever, and occurring at least 30 days after a previous attack. This definition has previously been applied in an MSBase relapse phenotype analysis.\textsuperscript{36} The follow-up period for each eligible patient across which relapse events
could be observed was defined as the period spanning the date of first EDSS assessment through to the date of the most recent EDSS assessment recorded in the registry prior to the data of data extract and compilation. In instances where the exact day of relapse onset was unavailable or unable to be determined for a particular month, clinics used either the 1<sup>st</sup> or 15<sup>th</sup> the day of the month as default dates. Of the 32,762 relapses analysed in this report, 7913 (24.2%) and 4594 (14.0%) were recorded on the 1<sup>st</sup> and 15<sup>th</sup> day of the month respectively, significantly higher than the proportions recorded on any other day of the month which ranged from 0.8% through 5.6%. To correct for this, relapses recorded on either the 1<sup>st</sup> of 15<sup>th</sup> day of the month were randomized to a day within a 15 day interval either side of both these default dates. The internal validity of the this approach was confirmed via sensitivity analyses which demonstrated that the modelled estimate of peak relapse date under default date randomization was not significantly different from a model using either the original reported dates or excluding default dates entirely.

**Ultraviolet radiation**

Daily average erythemally-weighted ambient UVR for each month from 1979 to 2004 inclusive was sourced from the National Aeronautics and Space Administration Earth Probe Total Ozone Mapping Spectrometer for all individual locations included in the analysis.  

**Protocol Text:**

NOTE: Each step described corresponds to a section of Stata code with the same number in the code file provided. Stata command names have been italicised in the following protocol.

1) Prepare and plot the observed relapse onset data

1.1) Open a do file by clicking on the “New Do-file Editor” button and use the *generate* command to calculate the number of relapse onsets dated to each of the twelve calendar months for each of the
three geographic levels to be modelled: location, hemisphere and global. Action command by clicking the “Execute (do)” do-file action button in the do-file.

1.2) Use the `sktest` command to test the underlying distribution of relapse counts for normality using a Shapiro-Wilk test. Select code and click “Execute (do)”.

NOTE: In the presence of significant skew, apply a natural log transformation and subsequently test the log-transformed relapse count variable for approximate normality by reapplying the Shapiro-Wilk test.  

1.3) Use the `generate` command to create a new variable “north_month” for southern hemisphere calendar months offset by +6 to allow plotting of both northern and southern hemisphere relapses by season along the same horizontal axis. Select code and click “Execute (do)”.

1.3.1) Graph a scatterplot of observed monthly relapse onsets with relapse frequency on the y-axis and calendar month on the x-axis for each hemisphere using the `twoway scatter` command. Repeat for each location. Observe pattern of peaks and troughs in relapse onset over the calendar year by viewing each plot in the graph viewer the automatically opens to screen.

1.4) Use the `radar` command to draw radar plots of the distribution of relapse frequency by calendar month with each radar axis capturing a single month ordered in a clockwise manner. Select code and click “Execute (do)”.

1.4.1) Repeat for all sites. Observe pattern of peaks and troughs in relapse onset over the calendar year by viewing each plot in the graph viewer the automatically opens to screen.
1.5) Run the `seast` command to apply an Edward’s test of seasonality across the observed relapse data.\(^{39,40}\) Repeat for all geographic levels.

2) **Model building and selection**

2.1) Use the `generate` command to specify the annual cycle sine and cosine trigonometric functions to be used in the regression. Select code and click “Execute (do)”.

2.2) Use the `regress` command specify the form of the base model with relapse count as the dependent outcome variable and the sine and cosine terms calculated in step 2.1 as the primary explanatory variables.

2.2.1) Add location-specific UVR to the base model as an additional adjusting covariate and use the analytic weight `aweight` option to weight the model for the number of patients contributed by each location. Select code and click “Execute (do)”.

NOTE: Record the model coefficient of determination (R\(^2\)) and the residual error in the results window that automatically opens to screen.

2.3) Store the model predicted monthly log(relapse) using the `predict` command. Convert log relapses back to integer relapse counts by exponentiating the log(relapse) term using the `generate` command. Select code and click “Execute (do)”\(^{,40}\). Repeat for all sites.

2.4) Overlay the exponentiated predicted monthly relapse estimates from 2.3 over the observed monthly relapse data using the `twoway scatter` command. Select code and click “Execute (do)”\(^{,40}\).

2.4.1 Repeat for all sites. View each plot in the graph viewer.
2.5) Use the `regress` command to expand the model specified in 2.2 by adding an additional harmonic sine/cosine pair. Select code and click “Execute (do)”.

NOTE: Record the residual error and the coefficient of determination. Save and transform model estimates as per 2.3 and plot model estimates over observed data as per 2.4. Repeat for all sites.

2.6) Use the `regress` command to further expand the model specified in 2.2 by adding two additional harmonic sine/cosine pairs. Select code and click “Execute (do)”.

NOTE: Record the residuals and the coefficient of determination. Compare this model directly with the base model using a likelihood ratio test. Save and transform model estimates as per 2.3 and plot model estimates over observed data as per 2.4. Repeat for all geographic levels.

3) Estimating peak relapse probability

3.1) Use the non-linear combination of estimators function (`nlcom`) to calculate the point estimate and 95% confidence interval for the phase-shift, using the best-fitting model identified from steps 2.1 through 2.6. Select code and click “Execute (do)”.

3.1.1) Convert these point estimates and associated confidence intervals to numbers representing calendar dates of peak relapse frequency ($T_{\text{max}}$) and trough relapse frequency ($T_{\text{min}}$), where $1$=1$^{\text{st}}$ January and $365$=31$^{\text{st}}$ December and $T_{\text{max}} = \text{phase-shift} + (365/4)$ and $T_{\text{min}} = \text{phase-shift} + ((365/4)*3)$. Repeat for all geographic levels.
3.2) Use the *generate* command to calculate peak-to-trough difference \((T_{\text{max}} - T_{\text{min}})\) for each location, standardized for every 100 patients per site. Use a Wilcoxon rank-sum test to compare standardized peak-to-trough difference by latitude range. Select code and click “Execute (do)”. 

4) **Modelling ultraviolet radiation data**

4.1) Run *use* command to load the UVR data. Calculate median monthly UVR for each location using the *egen* command. Select code and click “Execute (do)”.

4.2) Graph a scatterplot of monthly UVR (y-axis) by calendar month (x-axis) for each location using the *twoway scatter* function. View each plot in the graph viewer the automatically opens to screen.

4.3) Repeat step 1.2 for the UVR data and use the *regress* command to specify a base model of location-level annual UVR trend where monthly UVR is specified as the dependent outcome variables and the sine and cosine trigonometric functions specified in step 2.1 are incorporated into the model as the explanatory variables.

4.4) Repeat steps 2.4 through 2.6 for the UVR model and limited to the location-specific models only.

4.5) Using the best-fitting model of location-specific monthly UVR identified in steps 4.2 through 4.4 use the *generate* command to calculate the phase-shift point estimate and associated 95% confidence interval for UVR by again applying the double-angle formulae specified in step 3.1. Calculate \(T_{\text{min}}\) (date of trough UVR) for each location using the formula described in step 3.1. Select code and click “Execute (do)”.

5) **Modelling UVR-trough-to-relapse-peak lag**
5.1) Append the model-estimated dates of seasonal UVR trough from step 4.5 and relapse peak dates from step 3.1 for each location using the `merge` command. Use the `generate` command to calculate the time lapsed in months between UVR trough date and subsequent relapse peak date. Select code and click “Execute (do)”. 

5.2) Use the `sktest` command to test the UVR-trough-to-relapse-peak lag variable for significant departures from normality using a Shapiro-Wilk test. Select code and click “Execute (do)”. 

5.3) Append location-level latitude data to dataset using the `merge` command. Convert relative latitude to absolute latitude using the `abs(x)` function. Select code and click “Execute (do)”. 

5.4) Using the `regress` command, test the linearity of the relationship between lag and absolute latitude by running both linear and quadratic regressions and comparing residuals. Select code and click “Execute (do)”. 

5.5) Using `regress`, specify a linear mean regression model with UVR-trough-to-relapse-peak lag as the dependent outcome variable and absolute latitude in units of 10 degrees as the predictor variable. Weight the model for the number of patients contributed by each location using the `aweights` regress option. Select code and click “Execute (do)”. 

5.6) Use the `twoway scatter` command to plot absolute latitude on the y-axis against UVR-trough-to-relapse-lag in months on the x-axis. Overlay a line of best fit using `lfit` graph option. Visualise the relative patient weights of each location using the `aweight` analytical weights option. Select code and click “Execute (do)”. 

6: Sensitivity analyses of patient-level relapse propensities
6.1) Use the `mepossion` command to specify a mixed-effects Poisson regression where monthly relapse count is the dependent outcome variable, the sine and cosine trigonometric functions specified in step 2.1 are again incorporated into the model as the fixed variables, baseline EDSS, age at MS onset and prior exposure to MS specific disease-modifying treatment are included as potential confounders and unique patient identifier is specified as a random effect. Select code and click “Execute (do)”.

6.2) Repeat steps 2.4 through 2.6 to identify the best-fitting Poisson model.

6.3) Repeat step 3.1 over the Poisson model to calculate the date of peak relapse frequency. Compare results with the primary analysis.

6.4) Use the `generate` command to recalculate UVR-trough-to-relapse-peak lag in months for each location as described in step 5.1, using the patient-level Poisson model estimates of peak relapse date derived in step 6.3. Select code and click “Execute (do)”.

6.5) Use the `regress` command to remodel absolute latitude as a predictor of lag as described in step 5.5 and compare results with the primary analysis. Select code and click “Execute (do)”.

**REPRESENTATIVE RESULTS:**

The application of trigonometric regression to 32,762 relapse events sourced from 46 clinical centers across 20 countries was the basis for providing a defensible statistical argument for the observation that the timing of relapse onset in MS is cyclic and seasonal across both hemispheres and that the duration between seasonal UVR trough and subsequent relapse peak correlates with latitude. Critical to this was the reliance upon plot analysis to guide the necessarily iterative process of model development, evaluation and refinement.
Analysis of relapse frequency by calendar month scatter plots of the observed data suggested an annual cycle with a spring peak and autumn trough across all geographical levels. Radar plots of the global relapse data confirmed northern hemispheric relapses peaked in May (Figure 1a). This spring peak persisted when southern hemisphere relapse onset data was combined with the northern data (Figure 1b), with southern locations demonstrating a November peak. Autumnal troughs were also recorded in both hemispheres with the lowest frequency of northern and southern hemispheric relapses observed in November and May respectively. An Edward’s test further confirmed that relapse onset demonstrated significant departures from a uniform, non-seasonal distribution. Taken together, these results suggested that the periodic temporal variation observed in MS relapse onset at all three levels of geography best described a single annual cycle consisting of a single peak and a single trough separated by a regular six month interval. Thus a trigonometric regression model specified with a single pair of sine and a cosine functions was selected as the base case model across both hemispheres (Figure 2). When compared to competing trigonometric model solutions expanded to include two or three period harmonics, the base model across northern hemispheric locations minimised the residual square error and returned a superior fit of the observed data (p<0.0001, adjusted $R^2 = 0.263$) when compared to either a model incorporating an additional harmonic (p=0.0001, adjusted $R^2 = 0.198$) or an additional two harmonics (p=0.0014, adjusted $R^2 = 0.181$). Similarly the same base model out-performed the extended-harmonic alternatives when applied to the southern hemisphere with the base model (p<0.0001, adjusted $R^2 = 0.241$) again minimizing the residual differences between the observed and estimated data relative to the model incorporating two additional harmonics (p<0.0001, adjusted $R^2 = 0.167$); the one-additional harmonic model described a similar fit relative to the base (p<0.0001, adjusted $R^2 = 0.243$).

Importantly for the modelling of location-specific latitude as a predictor of UVR-trough-to-relapse-peak lag, the base model again out-performed either of the extended-harmonic models at the level of the individual geographical locations.
Figure 1: Radar plots of observed global relapse frequency by month in the 1a) northern hemisphere, 1b) combined northern and southern hemispheres
Figure 2: Plot comparing observed monthly relapses by hemisphere with predicted relapses using the base-case geometric model describing a single annual cycle of one peak and one trough separated by six months.

Using the base model specified on a single sine/cosine pair, the phase-shift across all relapses globally was estimated at -24.8 (95% CI -45.8, -3.9), translating into an estimated northern hemispheric peak relapse onset date of the 7th March (95% CI: 10th February, 28th March) and a southern hemispheric peak date of the 5th September (95% CI: 10th August, 26th September). There was no difference in the phase-shift estimate by hemisphere (test of interaction: p=0.254). Mean (SD) standardized peak-to-trough relapse difference was 7.6 (6.6) relapses per 100 patients. Although centers located at an absolute latitude of 40 degrees or more recorded a larger peak-to-
tough difference (mean 8.6, SD 7.6) relative to sites located within an absolute latitude range of 20 through 39 degrees (mean 5.7, SD 3.3), this difference was not statistically significant (p=0.135).

Scatterplot analysis of UVR by calendar month suggested that the base model defined on a single sine/cosine harmonic pair as described above was similarly appropriate for UVR seasonality, at all geographical levels. As illustration, Figure 3 depicts the regression modelled monthly UVR estimates overlaid on the observed UVR data for four selected individual locations, two from each hemisphere. What can be appreciated from these plots is just how closely the modelled estimates, based on an annual cycle single peak and trough sine regression, confers to the observed data. The base UVR model again outperformed either of the extended-harmonic models in terms of minimizing residuals and a superior coefficient of determination.

Figure 3: Plots comparing observed median monthly UVR with base model predicted UVR for Montreal, Canada; Melbourne, Australia; Bari, Italy & Buenos Aires, Argentina.
Overlaying the cyclical UVR sinusoid curve over the equivalent curve for relapse onset suggested that trough UVR consistently preceded peak relapse onset probability. Furthermore this lag appeared to shrink the further north or south a particular location was sited away from the equator. Applying a linear regression of the mean, every 10 degrees of latitude away from the equator in either hemisphere was associated with a statistically significant decrease in this lag of 28.5 days in the UVR-trough-to-relapse-peak lag (95% CI: 3.29, 53.7; p=0.028). As Figure 4 demonstrates, as absolute latitude increased away from the equator in both hemispheres, the sooner relapses peaked following the winter UVR trough. There was no difference in this association by hemisphere (test of interaction p=0.811).

The patient-level mixed effects Poisson extension of the primary trigonometric sine regression returned very similar results with a peak relapse date estimated at just two days later than that estimated by the primary base model (9th March compared with the 7th March for northern hemisphere locations, 7th September versus 5th September for southern locations). Similarly the UVR-trough-to-relapse peak lag was comparable under either the primary or sensitivity models, with the patient-level Poisson extension demonstrating a mean only 4.1 days different in lag (mean lag =24.8 days, 95% CI 2.0, 49.2) relative to the primary location-level model. Again, there was no difference in this association by hemisphere (test of interaction, p=0.671).
**DISCUSSION:**

The protocol described herein details a systematic regression based technique, guided by visual plot analysis, of global MS relapse onset data. It takes as a starting point a relatively simple descriptive analysis of relapse data from 20 countries across both hemispheres, allowing the user to explore theories regarding the temporality of relapse onset timing in MS and testing these theories formally through the use of trigonometric models. Through a step-wise process of first plotting global relapse onset data and then systematically graphing and assessing candidate geometric fits of the observed data, a latitude-dependent correlation between seasonal trough UVR and subsequent peak relapse onset probability was observed, a correlation hitherto unprecedented in MS epidemiology. Furthermore by combining trend visualization with formal statistical modelling, this analysis also confirmed prior meta-analysis suggesting seasonality was a factor in relapse onset timing in the
northern hemisphere and, also for the first time, extended this observation to the southern hemisphere.

Trigonometric regression modelling is a flexible tool for formally exploring cyclical, time- or season-dependent periodic phenomena, permitting statistical characterisation of trend data that conforms to geometric shapes such as the annual cyclical sinusoid curve observed in both the relapse onset timing and UVR data explored in this report. However given the range of shapes and structures that complicated, multi-factorial epidemiological trend phenomena such as relapse onset timing may potentially assume, visualization of both the original data and the differences between such observed data and those predicted by a particular model (i.e. the residuals) are critical for both the hypothesis-generating (relapse onset timing varies seasonally across the year) and hypothesis-testing phases of this study (this seasonality is predictable and best described using a sine regression). The result is a suite of novel, empirically-grounded inferences regarding the potential global influence of season and latitude in disease exacerbation patterns in MS.

The critical step within the protocol was perhaps the simplest to execute, the visualization of the observed relapse onset data using simple descriptive scatterplots. Given the multitude and diversity of possible temporal structures periodic data may take, simple graphs of observed data provide both the an empirical basis for forming a hypothesis around relapse onset patterns as well as the starting point for building models that best capture these trends and which can subsequently be used for statistical inference and prediction. A key modification engineered into the protocol was the systematic comparison of the base model against alternative models incorporating additional trigonometric harmonic functions. “Best” fit is a relative state and only by testing the performance of the base model against plausible alternatives was best fit in this case able to be determined. The other key step was replicating each of the relapse probability and UVR models at all three distinct levels of geography – global, hemisphere and location. Not only did this provide internal validation
of the primary results (the higher powered trends observed at the global and hemispheric levels were replicated at the location level) it also permitted troubleshooting of the code used to run the plots and models. Unexpected results or implausible model fits, not always evident at the level of the global or hemispheric analysis, derived at the level of the location were used as a red flag for quality checking the code used over all levels of geography. This provided confidence that the seasonal cycles and latitude patterns observed globally were not an artefact of data aggregation or miscoding. A further advantage of this protocol is that not only can it capture and describe seasonality and the influence of hemisphere and latitude with an appropriate robustness, it also adjusts these associations for potential confounding from patient-level propensities for relapse including differing levels of disability and varying disease-modifying drug exposure prior to relapse. This allows us to better isolate season and latitude as independent predictors of relapse probability resulting in estimates of effect that better approximate the truth. This is particularly important given the potential clinical consequences of this research.

The observation of a predictable, latitude-dependent lag between winter trough UVR levels and subsequent relapse peak frequency may in part relate to an influence of changing vitamin D status at a given geographical location, each with its own unique UVR profile. Several vitamin D-mediated immunomodulatory correlates with MS relapse onset probability have previously been observed including shifting T helper lymphocytes away from a pro-inflammatory Th1 profile to the less inflammatory Th2\textsuperscript{31,34} and inhibition of dendritic cells and IgM/IgG antibody production\textsuperscript{31,35-38}. Coupling this to the observation of a potential role for both season and latitude in the kinetics of relapse timing, this suggests a role in clinical practice for latitude-specific, location-appropriate vitamin D supplementation for reducing the probability of future relapse. Of course despite this suggestion, the MSBase study did not collect longitudinal data on patient-level vitamin D status nor formal UVR skin exposure quantification and thus this theorized inverse correlation between vitamin D status and subsequent relapse probability remains exactly that, a hypothesis only. Formal,
appropriately powered randomized clinical trials are required to establish causality. Two such trials of vitamin D monotherapy, the Australian/New Zealand PREVANZ trial (registration ACTRN12612001160820) and the French “D-lay MS” study (registration PHRC-N/2012/ET), are currently in progress.

Perhaps most notably, this study is illustrative of the possible synergy available to epidemiologists from the combination of formal statistical modelling and diagnostics with data visualization techniques. The significance of this technique relative to other forms of time-series analyses lies with its rejection of the assumption of conventional time series analyses that any underlying temporality is predominantly a random process. By comparison trigonometric regression explicitly seeks out structures in the temporal variation of cyclic, periodic phenomena such as MS relapse. As such trigonometric models are exquisitely reliant upon systematic visualization of both observed data and modelled estimates to guide and corroborate the model building and evaluation process, every step of the way. Neither the visualization or the modelling would have been sufficient in isolation – plot analysis was necessary for establishing realistic hypotheses regarding the presence and structural form of seasonal and latitudinal influences of relapse probability and then testing the performance of the resultant models whilst trigonometric regression was necessary for both quantifying these relationships, adjusting for important confounders, and providing a measure of certainty as to how plausible these associations are.

The technique described herein is a powerful method for isolating the role or influence of seasonality or latitude on complex, multifactorial events such as the timing of MS relapse. As such it has future potential for wide application for studying other clinical or biological phenomena which are known or suspected of varying systematically with season and/or latitude. This technique would be particularly relevant for prediction in disease epidemiology, both in terms of communicable and non-communicable disease where the timing of key events such as an infection or disease
progression are complex and often driven by a multitude of both environmental factors (season, temperature, latitude) and patient-level characteristics (age, comorbidities, exposure to treatment). Such a tool may assist in risk-stratification of patients more likely to experience an adverse health event and thus guide earlier interventions.

ACKNOWLEDGEMENTS:
The authors would like to thank Ivan Hanigan for his support in extracting and interpreting the ultraviolet radiation satellite data. The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. RL is supported by a NHMRC Career Development Award [ID 1004898].

DISCLOSURES:
Tim Spelman received honoraria for consultancy and funding for travel from Biogen Idec Inc; Orla Gray received travel support from Biogen Idec, Merck Serono and Novartis; compensation for serving on scientific advisory boards from Biogen Idec, Genzyme, Novartis and Merck Serono; Robyn Lucas did not disclose any competing interests and Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.

REFERENCES:


C.1.1.1 – Protocol code

*SECTION 1: PREPARE AND PLOT OBSERVED RELAPSE ONSET DATA

*1.1) Calculate relapses per month - by location, hemisphere & global

generate count=1
collapse (count) nrelapse_location = count, by( calendar_month location_flag)
collapse (count) nrelapse_hemisphere = count, by(calendar_month hemisphere_flag)
collapse (count) nrelapse_global = count, by(calendar_month)

*1.2) Test cumulative relapses per calendar month for skew

sktest nrelapse_location
sktest nrelapse_hemisphere
sktest nrelapse_global

generate log_nrelapse_location=log(nrelapse_location)
generate log_nrelapse_hemisphere=log(nrelapse_hemisphere)
generate log_nrelapse_global=log(nrelapse_global)

sktest log_nrelapse_location
sktest log_nrelapse_hemisphere
sktest log_nrelapse_global

*1.3) Calculate month offsets
generate north_month = cond(hemisphere_flag == "northern", month_test, month_test +6 )
replace north_month = north_month -12 if north_month > 12

*1.3.1) Scatterplot of observed monthly relapse data by season – example provided is for hemisphere

twoway scatter nrelapse_hemisphere north_month,
    by(hemisphere_flag, yrescale col(1) note("Observed relapses") ) ||
xlabel( 2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")
xline(3.5 6.5 9.5 ) xtitle("Season")
xtitle("Season")
ylabel(, angle(0))
legend( label(1 "Observed relapses"))

*1.4) Generate radar plot of monthly relapse data

*LOCATION
radar north_month nrelapse_location if location_flag=="Insert Location Here"

*HEMISPHERE
radar north_month nrelapse_hemisphere if hemisphere_flag=="Northern"
radar north_month nrelapse_hemisphere if hemisphere_flag=="Southern"

* GLOBAL
radar north_mon nrelapse_global
*1.5) Edward’s test for seasonality (seasonal departures from a uniform distribution)

seast nrelapse_location, edwards sector(calendar_month)
seast nrelapse_hemisphere, edwards sector(calendar_month)
seast nrelapse_global, edwards sector(calendar_month)

*SECTION 2: MODEL BUILDING AND SELECTION (Example = global)

*2.1) Specify the annual cycle trigonometric function to be used in the models

generate theta = 2*3.14159*( north_month - 0.5)/12
generate sin1 = sin(theta)
generate cos1 = cos(theta)

*2.2) Specify the base model

xi: regress log_nrelapse_global sin1 cos1 location_UVR [awi=patient_count]

*2.3) Save model-estimated monthly relapse and exponentiate log term

predict log_prednrelapse_global
generate prednrelapse_global = exp(log_prednrelapse_global)
2.4) Re-plot observed relapse onset data and overlay model estimates

```
twoway scatter prednrelapse_global north_month,  
    by(hemi, yrescale col(1) note("Model: <Insert geography here>")) || 
    conn prednrelapse_global north_month, conn(L)  
    /* xlabel(1(1)12,labsize(small)) */  
    xlabel(2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")  
    xline(3.5 6.5 9.5 ) xtitle("Season")  
    ytitle("Season")  
    ylabel(, angle(0))  
    legend( label(1 "Observed relapses") label(2 "Predicted relapses"))
```

2.5) Comparator model – model 2

```
xi: regress log_nrelapse_global sin1 cos1 sin2 cos2 location_UVR  
    [awei=patient_count]  
predict model2_log_prednrelapse_global  
gen model2_prednrelapse_global = exp(model2_log_prednrelapse_global)  
twoway scatter model2_prednrelapse_global north_month,  
    by(hemi, yrescale col(1) note("Model: <Insert geography here>")) || 
    conn prednrelapse_global north_month, conn(L)  
    /* xlabel(1(1)12,labsize(small)) */  
    xlabel(2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")  
    xline(3.5 6.5 9.5 ) xtitle("Season")  
    ytitle("Season")  
    ylabel(, angle(0))  
    legend( label(1 "Observed relapses") label(2 "Predicted relapses"))
```
*2.6) Comparator model – model 3

\[ \text{xi: regress log}_n\text{relapse}_\text{global} \sin 1 \cos 1 \sin 2 \cos 2 \sin 3 \cos 3 \text{ location_UVR} \]
\[ \text{[awei=} \text{patient}\_\text{count}] \]

\[ \text{predict model3}_\log\text{prednrelapse}_\text{global} \]

\[ \text{gen model3}_\text{prednrelapse}_\text{global} = \exp(\text{model3}_\log\text{prednrelapse}_\text{global}) \]

\[ \text{twoway scatter model3}_\text{prednrelapse}_\text{global} \text{ north}_\text{month}, \]
\[ \text{by(hemi, yrescale col(1) note("Model: <Insert geography here>")) ||} \]
\[ \text{conn prednrelapse}_\text{global} \text{ north}\_\text{month, conn(L)} \]
\[ /* \text{xlabel(1(1)12,labsize(small)) */} \]
\[ \text{xlabel( 2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")} \]
\[ \text{xline(3.5 6.5 9.5 ) xtitle("Season")} \]
\[ \text{xtitle("Season")} \]
\[ \text{ylabel(, angle(0))} \]
\[ \text{legend( label(1 "Observed relapses") label(2 "Predicted relapses")})} \]

*SECTION 3: ESTIMATING PEAK RELAPSE PROBABILITY AND PEAK-TO-TRough DIFFERENCE

3.1) Calculate phase shift, T_max and T_min

\[ \text{nlcom (phase}\_\text{shift: } \text{atan(}_b\cos 1/}_b\sin 1)*365.25/(2\times3.14159) ) \]

generate relapse_tmax = phase_shift + (365/4)
generate relapse_tmin = phase_shift + ((365/4)*3)

3.2) Calculate peak-to-trough difference

\[ \text{by location}\_\text{flag: gen peak}\_\text{trough}\_\text{difference }= \text{relapse}\_\text{tmax} - \text{relapse}\_\text{tmin} \]
\[ \text{by location}\_\text{flag: gen standardized}\_\text{peak}\_\text{trough}\_\text{difference} = (peak\_\text{trough}\_\text{difference/patient}\_\text{count})\times100 \]
generate  absolute_latitude_range="20_39" if  absolute_latitude>=20 & absolute_latitude<40
replace absolute_latitude_range="40plus" if absolute_latitude>=40
ranksum standardized_peak_trough_difference, by(absolute_latitude_range)

*SECTION 4: MODELLING UVR DATA

*4.1) Calculate mean monthly UVR for each location

by location_flag calendar_month: egen monthly_UVR=median(UVR)

*4.2) Scatterplot – mean monthly UVR by calendar month

twoway scatter  monthly_UVR month,
   by(city, yrescale col(1) note("Model: insert location here")) ||
xlabel( 2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")
xline(3.5 6.5 9.5) xtitle("Season")
ylabel(,angle(0))
legend( label(1 "Observed UV") )

*4.3) Specify base UVR model

sktest monthly_UVR
generate log_monthly_UVR = log(monthly_UVR)
genenate theta = 2*3.14159*( month - 0.5)/12
generate sin1 = sin(theta)
generate cos1 = cos(theta)
x: regress log_monthly_UVR sin1 cos1
predict predicted_log_monthly_UVR
generate predicted_monthly_UVR = exp(predicted_log_monthly_UVR)

twoway scatter monthly_UVR month,
   by(city, yrescale col(1) note("Model: <insert location here>")) ||
   conn predicted_monthly_UVR month, conn(L)
xlabel(2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")
xline(3.5 6.5 9.5) xtitle("Season")
ylabel(), angle(0)
legend(1 "Observed UVR" 2 "Predicted UVR")

*4.4) Specify and test extended alternate models for UVR

xi: regress log_monthly_UVR sin1 cos1 sin2 cos2
predict model2_predicted_log_monthly_UVR

generate model2_predicted_monthly_UVR = exp(model2_predicted_log_monthly_UVR)

twoway scatter monthly_UVR month,
   by(city, yrescale col(1) note("Model: <insert location here>")) ||
   conn model2_predicted_monthly_UVR month, conn(L)
xlabel(2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")
xline(3.5 6.5 9.5) xtitle("Season")
ylabel(), angle(0)
legend(1 "Observed UVR" 2 "Predicted UVR")

xi: regress log_monthly_UVR sin1 cos1 sin2 cos2 sin3 cos3
predict model3_predicted_log_monthly_UVR

generate model3_predicted_monthly_UVR = exp(model3_predicted_log_monthly_UVR)
twoway scatter monthly_UVR month, 
    by(city, yrescale col(1) note("Model: <insert location here>")) ||
    conn model3_predicted_monthly_UVR month, conn(L)
xlabel( 2 "Winter" 5 "Spring" 8 "Summer" 11 "Autumn")
xline(3.5 6.5 9.5) xtitle("Season")
ylabel(,angle(0))
legend( label(1 "Observed UVR") label(2 "Predicted UVR"))

4.5) Calculate phase shift, Tmax and Tmin

nlcom (UVR_phase_shift: atan(_b[cos1]/_b[sin1])*365.25/(2*3.14159) )
generate UVR_tmax = UVR_phase_shift + (365/4)
generate UVR_tmin = UVR_phase_shift + ((365/4)*3)

*SECTION 5: MODELLING UVR-THROUGH-TO-RELAPSE PEAK LAG

5.1) Calculate lag in months

merge location using "file_path\UVR_and_relapse_dates.data"
generate lag_months=(relapse_tmax - UVR_tmin)/30.42

5.2) Test lag for skew

sktest lag_months

5.3) Convert relative latitude to absolute latitude in units of 10 degrees
generate absolute_latitude = abs(latitude)
generate absolute_latitude_10 = absolute_latitude/10

5.4) Run linear and quadratic trend models

regress lag_months calendar_month
regress lag_months calendar_month calendar_month^2

5.5) Model absolute latitude as a predictor of lag

regress lag_months absolute_latitude_10 [aweight=patient_count]

5.6) Plot latitude by lag

twoway (scatter absolute_latitude lag_months [aweight=patient_count] )
(lfit absolute_latitude lag_months [aweight=patient_count] )

*SECTION 6: SENSITIVITY ANALYSIS - PATIENT-LEVEL RELAPSE PROPENSITIES
(EXAMPLE = GLOBAL)

6.1) Specify patient-level mixed-effects Poisson model

xi: mepoisson nrelapse_global sin1 cos1 location_UVR baseline_edss
    age_MS_onset treatment_exposure || patient_id
6.2) Run extended alternate models

\[ \text{xi: mepoisson nrelapse\_global sin1 cos1 sin2 cos2 location\_UVR} \]
\[ \text{baseline\_edss age\_MS\_onset treatment\_exposure || patient\_id} \]

\[ \text{xi: mepoisson nrelapse\_global sin1 cos1 sin2 cos2 sin3 cos3 location\_UVR} \]
\[ \text{baseline\_edss age\_MS\_onset treatment\_exposure || patient\_id} \]

6.3) Re-calculate \( t_{\text{max}} \)

\[ \text{nlcom (phase\_shift: atan(_b[cos1]/_b[sin1])*365.25/(2*3.14159))} \]
\[ \text{generate SA\_relapse\_tmax = phase\_shift + (365/4)} \]
\[ \text{generate SA\_relapse\_tmin = phase\_shift + ((365/4)*3)} \]

6.4) Re-calculate lag

\[ \text{generate SA\_lag\_months=(SA\_relapse\_tmax – UVR\_tmin)/30.42} \]

6.5) Remodel absolute latitude as a predictor of lag

\[ \text{regress SA\_lag\_months absolute\_latitude\_10 [aweight= patient\_count]} \]
C.2 Quantifying risk of early relapse in high risk patients with first demyelinating events - methods

C.2.1 Background

Nomograms are graphical representations of multivariate regression models and can be used in clinical practice to estimate individualized risk of later events such as conversion from clinically isolated syndrome (CIS) to clinically definite multiple sclerosis. Whilst there is a reasonable body of evidence supporting their use in oncology, particularly with regards to matching a baseline risk with subsequent mortality or post-remission recurrence,\textsuperscript{1-5} to date, there has only been limited cross-over of prognostic nomograms into non-oncological practice, despite multivariate time-to-event regression, the model represented by the nomogram, being a frequently employed method for studying time-to-event data.

C.2.1 Nomogram construction

A series of prognostic nomograms for time to conversion at baseline were derived using the effect size Hazard Ratios (HR) from the adjusted, multivariate models as inputs. The nomograms presented are thus graphical representation of multivariate regression models. As described in section E, the suite of prognostic factors ultimately used in the nomograms were identified through multivariate time-to-event Cox PH modelling. The Cox PH model estimates the magnitude, direction and significance of the association between a candidate prognostic factors at CIS and the primary conversion end-point by modelling the instantaneous failure rate as, explicitly, a function of time (i.e. the hazard).\textsuperscript{2}

For example, the average hazard for experiencing a conversion event at, say, 12 months post CIS under a proportional hazards model parameterized as a function of age at CIS, sex, baseline disability (EDSS score) and neuroanatomical location of first symptoms is given by:
Conversion hazard, \( t \) = baseline hazard, \( t \) + \( \exp(\beta_{\text{age}} \cdot \text{age} + \beta_{\text{sex}} \cdot \text{sex} + \beta_{\text{EDSS}} \cdot \text{EDSS} + \beta_{\text{location}} \cdot \text{location}) \) \( (1) \)

where

\( t = \text{time} = 12 \text{ months} \)

baseline hazard, \( t \) = 12-month conversion hazard when all independent, explanatory covariates take the value of zero

\( \beta = \text{the regression parameter coefficient} \)

The left side of the equation captures the estimated or predicted probability of, in this case, conversion from CIS to CDMS by 12 months post-CIS. It is this metric which is then translated into the key conversion probability quantity presented in the personalized risk nomograms. By using the effect sizes estimated by regression model as inputs and converting these to a point scale linked to an event probability, the nomogram is effectively a graphical representation of the multivariate Cox PH regression model.

All nomograms presented and described in section E were derived in R (R Foundation for Statistical Computing, Vienna, Austria) using the following example code:

\[ R> \text{require('rms')} \]
\[ R> \text{require('survival')} \]

This calls in the Regression Modelling Strategies (RMS) and Survival modules required to run the analysis. The following code calls the relevant data in, applies required formatting, fits the Cox model, store the X-year conversion estimates (where X=0.5, 1, 2, 3, 4, 5) and the standard error associated with each point estimate:
R> coxData <- read.csv("second_attack_dataset_abridged_extended
version_modifed_test_25may15.csv", header = TRUE)

# Add label for the variables
R> label(coxData$first_symptoms) <- 'First symptom location'
R> label(coxData$Infratentorial) <- 'T1 Gd+ Infratentorial Lesions'
R> label(coxData$Periventricular) <- 'T1 Gd+ Periventricular Lesions'

R> coxData$first_symptoms <- factor(coxData$first_symptoms,
levels = c("Brainstem", "Optic_pathway", "Spinal_cord", "Supratentorial"),
labels = c("Brainstem", "Optic Pathway", "Spinal Cord", "Supratentorial"))

R> coxData$Infratentorial <- factor(coxData$Infratentorial,
levels = c("0", "1"),
labels = c("0", "1+"))

R> coxData$Periventricular <- factor(coxData$Periventricular,
levels = c("0", "1", "2"),
labels = c("1-2", "0","3+"))

R> op <- datadist(coxData)
R> options(datadist = 'op')

R> fuSurv <- Surv(coxData$event_relapse_fu_years,
coxData$event_second_event_flag)
The next sequence of code takes the modelled effect sizes from the multivariate regression and converts this to a function for use in the nomogram translation:

```r
R> surv <- Survival(fucoxFit)
R> surv.mf <- function(x) {
    1 - surv(3, lp = x)
}
```

Finally, the next sequence of code calls the function that builds and graphs the nomogram:

```r
R> nomCox <- nomogram(fucoxFit, fun = surv.mf, funlabel = "Predicted Value", lp = TRUE)
R> plot(nomCox)
```

### C.2.2 Nomogram interpretation

Supplementary figure 1 of section E provides a worked example of how to use and interpret the conversion nomogram. The example provided is for a hypothetical, albeit typical and representative patient presenting at first symptoms (CIS) – female, aged 30 years at CIS onset, EDSS score of 2, supratentorial symptom location, 1-2 periventricular T2 hyperintense lesions and zero infratentorial lesions on cerebral MRI and Oligoclonal bands (OCB) detected on baseline CSF. In this particular
nomogram, female sex (the black arrow) matches to 8 points on the “Points” scale, an age at CIS of 30 years (red arrow) matches to 62 points and an EDSS of 2 (grey arrows) to 20 points. Similarly a supratentorial first symptoms location, 1-2 periventricular, zero infratentorial lesions and the presence of oligoclonal bands corresponds to, approximately, 60, 16, 0 and 40 points on the (upper) “Points” scale respectively, summing to a cumulative total of approximately a total of 138 points (refer blue arrow on the “Total Points” scale). Drawing a line down from the “Total Points” scale to the “Probability of 12-month conversion” scale demonstrates that 138 points corresponds to a 12-month conversion probability of 0.53 (or 53%) for this particular patient (green arrow).

As can be appreciated from this worked example, the nomogram, by using the independent hazard effects for each prognostic factor as estimated from the multivariate Cox PH model and expressing these as a number of points, in effect ranks the relative importance of each prognostic factor (and each constituent level or category within a factor) in predicting the conversion outcome scaled relative to the other covariates included in the predictive model used as input into the prognostic nomogram.

C.2.3 Nomogram validation

Validation of all nomograms was conducted through calculation of the concordance index and derivation of calibration plots.

C.2.3.1 Concordance index

The concordance index is a measure of nomogram accuracy and quantifies the overall magnitude of concordance between the nomogram-predicted and dataset observed probabilities of conversion to clinically definite MS. This c-index measures the proportion of convertor-non-convertor pairs where the convertor records a higher nomogram predicted probability of conversion at CIS than the non-convertor member of the pair. As such the c-index is analogous to the area under a receiver-
operator characteristic curve\textsuperscript{3,6} and forms a measure of the nomogram’s discrimination – the ability of the tool to correctly rank patients by actual conversion probability.

The various c-indices presented in section E were derived via a process of bootstrapped resampling. A series of 1500 bootstrapped random draws from the original analysis dataset of \( n=3296 \) MS patients with clinically isolated syndrome were performed and for each draw the nomogram-predicted 12-month conversion probability was recorded for both the convertor cohort and the non-convertors, from which the c-index was calculated. The application of bootstrapping estimation was used to permit relatively unbiased estimation of the indices. The following is an example of the code used to derive the c-index:

\[
R> \text{lp\_normalized} \leftarrow \text{svy\_cox\_fit}$x \%\% \text{as\_matrix(svy\_cox\_fit$coefficients)} - \\
+ \text{mean(svy\_cox\_fit}$x \%\% \text{as\_matrix(svy\_cox\_fit$coefficients)})
\]

\[
R> \text{cindex\_orig} \leftarrow \text{with(noNA, 1 - rcorr\_cens(lp\_normalized,} \\
+ \text{Surv(survival, surv\_cens))[[1]]}
\]

**C.2.3.2 Calibration**

Calibration plots provide a visual representation of the distance between the nomogram predicted conversion to CDMS probabilities and the actual conversion rates observed in the source datasets. By plotting nomogram prediction against the observed conversion probability, a calibration plot provides a graphical representation of the how well the nomogram works in terms of accuracy and discrimination by plotting the inbuilt error (predicted minus observed) of the prognostic nomogram. As further detailed in section E, the calibration involved grouping the patient sample according to their nomogram-predicted conversion probabilities and the means of these probability groups then compared against the empirically observed Kaplan-Meier conversion estimates on a calibration curve. In the various examples presented in section E, bootstrapping on 1500 random draws was
again employed to calibrate each nomogram. A bootstrapping approach to this part of the validation was chosen as it can prevent over-fitting of the observed data and thus, by extension, enable it to be applied to a separate patient sample."

R was again used to build and derive the calibration plots. An example of the code used is presented here and follows on from the R code presented above:

```r
R> calCox <- calibrate(fuCoxFit, cmethod = "KM", u = 2, m = 200, B = 1500, bw = FALSE,
                         type = "residual", sls = 0.05, what = "observed-predicted")
R> # Plot the graph.
R> plot(calCox, xlab = "Predicted X-year Survival",
       ylab = "Observed X-year Survival by the Kaplan-Meier method")
title(main = "Calibration graph of X-year outcome", cex.main = 1, font.main = 2)
R> plot(calCox, xlim = range(0:1), ylim = range(0:1),
       xlab = "Predicted X-year Survival",
       ylab = "Observed X-year Survival by the Kaplan-Meier method")
```

where B is the number of repetitions, u captures the conversion time of interest (6 months, 1/2/3/4/5 years) and m is the number of patients per group.

Using Figure 1 from section E as an example, the nomogram-predicted probability of, in this particular example, 12-month post CIS conversion is captured (represented) on the horizontal x-axis whilst the observed 12-month conversion probability is represented (captured) on the vertical y-axis.
The blue-grey diagonal line set at 45 degrees within the plot area represents perfect agreement between nomogram predicted and actual observed 12-month conversion probability. Points on either side of this diagonal represent deviations (differences) in (the) nomogram predicted conversion probability and the actual probabilities, with greater distance, either side, away from the diagonal representing a greater overall absolute error (abs(predicted minus observed))

C.2.4 References


C.3 Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment settings and products - methods

C.3.1 Background

Comparisons of study end-points between two or more treatment groups sourced from observational cohort data is challenging secondary to multiple sources of systematic differences in key confounder variables between treatment arms, introducing selection bias. This limits the investigator’s ability to attribute any observed difference in outcome between treatment arms to the treatment of interest in favour of selection effects. Propensity score matching (PSM) is a statistical technique for balancing two or more study groups of interest with respect to measured confounders.¹²

C.3.1.1 References


C.3.2 Binomial propensity score matching – 1:1 matching

The two-way treatment comparisons detailed in sections G employed a binomial PSM process using the following steps. For clarity, the switch to natalizumab vs switch to BRACE comparison detailed in section G.1 will be used as the example, however a similar process was employed for each of the two-way treatment vs comparator analyses presented in section E.

1) Assess the pre-matching imbalance

Paired tests were used to compare the distribution of key confounder factors at baseline.
2) Estimate the patient-level propensity score via logistic regression

The propensity score was derived from a logistic regression model, in which receipt of natalizumab was the outcome variable and the pre-treatment characteristics formed the explanatory variables. These baseline confounding factors were those characteristics available for the entire matched sample and included age, sex, disease duration, baseline EDSS, the number of pre-switch disease-modifying drug (DMD) initiations, the proportion of disease duration spent on DMD therapy, the number of pre-baseline DMD initiations as a proportion of disease duration, the total number of relapses within 12 and 24 months pre-baseline and the total number of steroid-treated relapses within 12 or 24 months prior to baseline (time at switch). An example of the code from Stata is provided here which employs Stata’s “psmatch2” program:

```
set seed 12345
tempvar sortorder
gen `sortorder' = runiform()
sort `sortorder'

xi: psmatch2 treatment_flag i.sex age_at_baseline_years disdurn_years dmtcnt propn_dmtdurn_disdurn propn_dmtcnt_disdurn_yrs bl_edss nrelap1y nrelap2y nrelap1ystrd nrelap2ystrd, logit noreplacement
```

This model derives a propensity score for each unmatched individual patient and represents the probability that a patient, regardless of whether they were sourced from the actual natalizumab group of the BRACE switch cohort, would have, theoretically, received or been assigned to the natalizumab group purely on the basis of the distribution of baseline factors included as explanatory variables in the propensity match. Subjects in the actual natalizumab arm were then matched to a comparable patient in the BRACE switch group on the basis of the similarity or closeness in their propensity scores. This process is then repeated until either the total pool of unmatched patients in
the smaller cohort is exhausted or the pre-set bounds or calipers around the range of acceptable differences in propensity score are reached.

The analysis dataset is then pruned to return just the matched pairs.

```
    gen pair = _id if _treated==0
    replace pair = _n1 if _treated==1
    rename pair matched_pair_ID
    sort matched_pair_ID
    keep if _weight==1
```

3) Check post-matching balance

The matched dataset is then analysed to assess whether the (now matched) treatment arms are in fact balanced with regards to the suite of confounder variables used in the match. This is undertaken via application of paired tests and the derivation of standardised differences, as further described below.

4) Estimate treatment effect in matched sample

Once the data is matched and the confounder balance confirmed, the primary analysis is now run on the matched data to estimate the treatment effect. This is described in detail within both the methods and in the various analyses presented in section E.

5) Sensitivity analysis

Whilst the PSM process ensures treatment arms are well balanced at baseline for the factors included in the derivation of the propensity score, it cannot adjust or otherwise control for imbalance in determinants or correlates of treatment assignment that were either not collected or incompletely observed (eg detailed MRI brain metrics). Thus a post-hoc sensitivity analysis is
important for estimating the minimum effect size required for an unobserved confounder to significantly alter the primary model results. Refer section C.3.6 below for further detail.

C.3.3 Binomial propensity score matching – many-to-one matching

The above example describes a 1:1 propensity match, where a single patient from the treatment arm is matched to a single patient from the comparator arm, to form a pair. Sections G.3, G.4 and G.5 extend this approach to apply a many-to-one PSM. Whilst the broad approach is largely the same as that described in the above, there are some minor, albeit important, distinctions.

The following is a sample of the R code used to affect a 2:1 match.¹²

First call in the relevant PSM package (source: http://gking.harvard.edu/matchit):

```r
R> require("MatchIt")
```

The call in the unmatched dataset and format as appropriate:

```r
R> patientMedicRaw <- read.csv("dataset_unmatched_13sep15_FTY.csv", header = TRUE)

R> patientMedicRaw$Row_no <- rep(1:dim(patientMedicRaw)[1])

R> col_idx <- grep("Row_no", names(patientMedicRaw))

patientMedicRaw <- patientMedicRaw[, c(col_idx, 1:ncol(patientMedicRaw))[-col_idx]]
```

Run the many-to-one PSM model:
Timothy Denis Spelman (58172)

```r
multiPSM <- matchit(formula = stay0_stop1_num_flag ~ gender + age_at_baseline_years +
                      disdurn_years + bl_edss + tot_relapse_12m_preBL +
                      tot_relapse_24m_preBL + tot_STRrelapse_12m_preBL +
                      tot_STRrelapse_24m_preBL + mcd_T1_1plus_match_use_only + mcd_T2_9plus_flag,
                      data = patientMedic, method = "nearest", distance = "logit",
                      ratio = 2, replace = FALSE)
```

```r
summary(multiPSM)
```

Extract and format the matched data:

```r
pairedData <- match.data(multiPSM)
```

```r
treatGroup <- match.data(multiPSM, group = "treat")
controlGroup <- match.data(multiPSM, group = "control")
```

```r
matchedID <- as.data.frame(multiPSM$match.matrix)
colnames(matchedID) <- c("Stay1", "Stay2")
```

```r
treatGroup$Stay1 <- as.character(matchedID$Stay1)
treatGroup$Stay2 <- as.character(matchedID$Stay2)
controlGroup$Row_no <- as.character(controlGroup$Row_no)
```

```r
for (i in 1:length(treatGroup$Row_no)) {
  for (j in 1:length(controlGroup$Row_no)) {
    if (treatGroup$Stay1[i] == controlGroup$Row_no[j]) {
      
    }
  }
}
```
```r
for (i in 1:length(treatGroup$Row_no)) {
  for (j in 1:length(controlGroup$Row_no)) {
    if (treatGroup$Stay2[i] == controlGroup$Row_no[j]) {
      treatGroup$Patient_STAY_2[i] <- 
        as.character(controlGroup$PATIENT_ID[j])
    }
  }
}

R> write.csv(treatGroup, file = "two-to-one matched pairs FTY_BRACE_13sep15.csv")
```

**C.3.3.1 References**

C.3.4 Multinomial propensity score matching across three treatment groups

Propensity score matching can be extended to match across more than two groups. Section G.6 describes an analysis across three propensity score matched treatment switch groups on a 1:1:1 basis. The method is very similar to that described above. However rather than employing a binomial logistic regression model to estimate patient-level propensity score, a generalised multinomial logistic approach is required.\textsuperscript{1,3} Under this generalised procedure, previously described by Rassen et al.,\textsuperscript{2} propensity score for each member of the three treatment groups is estimated via multinomial logistic regression. “Within-trio” optimised matching, without replacement, as utilised in the binomial matching described above, is then applied to identify triplets of patients, one from each of the three treatment arms, where the distance in propensity score across the triplet is minimised.

An alternate method, the “common-referent approach” was also trialled.\textsuperscript{2} This involves denoting one of the three treatment arms as the “referent” group – nominally labelled as treatment group 1. Patients from switch treatment group 2 were then matched to comparable group 1 patients. Similarly group 3 patients were likewise matched to group 1 subjects. Group 2 or 3 patients who matched to a common group 1 subject (the referent) were then identified.

A number of R modules were trialled for the n=1 matching including TriMatch (source: https://cran.r-project.org/web/packages/TriMatch/index.html) and the MNPS function from the TWANG package (source: https://cran.r-project.org/web/packages/twang/twang.pdf). An example of the relevant code using the TriMatch module is summarised here:

```r
R> library(TriMatch)
R> group <- read.csv("New dataset for propensity matching_18aug15.csv")
```
# Estimate propensity score;
R> group3PS <- trips(group, group$group_flag, formu=- gender + age_at_switch_years + disdurn_switch_years + count_12m_relapses + count_24m_relapses + bl_edss + dmtcnt + propn_dmtcnt_disdurn_yrs + propn_dmtdurn_disdurn + total_str_relapse_12m + total_str_relapse_24m)

# Create matched triplets;
R> tmatch2 <- trimatch(group3PS, nmatch=100, match.order = c("GROUP_1", "GROUP_3", "GROUP_2"))

# Create unique triplets
R> newmat <- OneToN(tmatch2, M1=1, M2=1)
R> newmat <- newmat[,c(1,2,3)]  ## remove distance columns
R> newmat <- data.frame(sapply(newmat, as.numeric))

# Split into each group for matching
R> groupNo <- group
R> groupNo$row_no <- seq(1, length(group$group_flag), by=1)
R> group2 <- groupNo[groupNo$group_flag == "GROUP_2", ]
R> group1 <- groupNo[groupNo$group_flag == "GROUP_1", ]
R> group3 <- groupNo[groupNo$group_flag == "GROUP_3", ]

# Merge data sets;
R> m1 <- merge(group2, newmat, by.x="row_no", by.y="GROUP_2")
R> for (i in 1:length(m1$row_no)) {
    for (j in 1:length(group1$row_no)) {
        if (m1$GROUP_1[i] == group1$row_no[j]) {
            m1$patient_G1[i] <- as.character(group1$PATIENT_ID[j])
        }
    }
}
R> for (i in 1:length(m1$row_no)) {
    # for (j in 1:length(group3$row_no)) {
    if (m1$GROUP_3[i] == group3$row_no[j]) {
        m1$patient_G3[i] <- as.character(group3$PATIENT_ID[j])
    }
    
    }
}

R> group3PSM <- m1[, -c(1, 15, 16)]

# Export
R> write.csv(group3PSM, file = "3 treatments PSM.csv")

C.3.4.1 References


C.3.5 Post-matching assessment of balance

The success of matching was assessed using two approaches; 1) paired tests and 2) derivation of standardised differences.

C.3.5.1 Paired tests

Wilcoxon rank-sum and chi-square tests were used to compare unmatched baseline characteristics by treatment arm as appropriate. Wilcoxon signed-rank and McNemar tests were used to compare baseline characteristics in the matched data for continuous variables and proportions, respectively.

C.3.5.2 Standardised differences

Standardized differences were calculated for both unmatched and matched comparisons, permitting direct comparison of different baseline characteristics with the same standardized units.

The standardised difference for a continuous explanatory variable (e.g. disease duration at baseline) is defined as:

\[
d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s^2_{treatment}}{2} + \frac{s^2_{control}}{2}}},
\]

where \(d\) denotes the standardised difference, \(\bar{x}_{treatment}\) and \(\bar{x}_{control}\) denote the sample mean of a matching covariate for the treatment (e.g. natalizumab switch) and comparator control arm (e.g. BRACE switch) respectively and \(s^2_{treatment}\) and \(s^2_{control}\) are that covariate’s sample variance in these same switch treatment arms.
The standardised difference for a binary, dichotomous explanatory covariate (e.g. gender) the formula for the standardised difference is defined as:

\[
\begin{align*}
    d &= \frac{(\hat{p}_{\text{treatment}} - \hat{p}_{\text{control}})}{\sqrt{\hat{p}_{\text{treatment}}(1 - \hat{p}_{\text{treatment}}) + \hat{p}_{\text{control}}(1 - \hat{p}_{\text{control}})}} \, .
\end{align*}
\]

where \( \hat{p}_{\text{treatment}} \) and \( \hat{p}_{\text{control}} \) are the prevalence of, for example, female sex in the natalizumab intervention and BRACE comparator arms respectively. A standardised difference of <0.2 is considered to represent good balance between matched groups.\(^1\)

### C.3.5.2.1 References


### C.3.6 Sensitivity analysis

A Rosenbaum sensitivity analysis can be used to estimate the resistance of the modelled treatment effect to unobserved confounding.\(^1,2\)

The R package “rbounds” was used for the various Rosenbaum sensitivity analyses described across chapters X through Y (source: [http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf](http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf)).\(^3\) The following summarises the key code items required to run both a Wilcoxon signed-rank Rosenbaum sensitivity (“psens”) and a Hodges-Lehman variant (“hlsens”):
The signed-rank approach provides Rosenbaum’s bounds around the p-value estimated by the primary
time-to-event regression model. The Hodges-Lehmann approach, by contrast, calculates
bounds for the additive treatment effect, rather than the p-value of the (same) treatment effect.
Whilst the former provides information on the minimum effect size required to change inference,
the Hodges-Lehmann bounds provide information on the difference in median effect sizes between
comparator groups.

C.3.6.1 References
C.4 Serial disability / time AUC plots

C.4.1 Background
Sections G.1, G.2 and G.4 present a novel serial disability/time area-under-the-curve analysis. Whilst the methodology is covered in part in the methods sections within each of the respective papers, the following is a full summary of the method. Whilst the summary uses the first-line natalizumab vs first-line BRACE comparison from section G.2 as a worked example, the same method was used in each of sections G.1, G.2 and G.4.

C.4.2 Deriving disability-time plots
Patients treated continuously with first-line natalizumab or BRACE therapy for a minimum of 24 months and recorded an EDSS assessment at both baseline (treatment initiation) and at 24 months (± 6 months) were eligible for the serial disability/time area under the curve (AUC) analysis.

For eligible patients from both treatment arms, all EDSS scores recorded within the 24-month treatment interval, including the baseline EDSS, were plotted. Cumulative serial disability was derived at the level of the individual patient by first calculating the total area under the curve between each pair of consecutive EDSS scores. This inter-EDSS area was defined on a rectangular rule as the product of the absolute difference in EDSS scores \(y_0, y_1, y_2, \ldots, y_n\) between consecutive assessment points multiplied by the time elapsed between consecutive assessment points \(t_0, t_1, t_2, \ldots, t_n\) for time points \(i = 1, \ldots, n-1\). These separate areas were then summed to derive the total 24-month cumulative AUC change from baseline \(\Delta \text{AUC}\) as described in formula 1.

\[
\Delta \text{AUC} = \sum (y_i - y_0)(t_{i+1} - t_i)
\]
Figure 1 provides an example of such a patient-level 24-month serial disability/time plot for a patient with a baseline EDSS of 2 at treatment initiation followed by five post-baseline EDSS assessment points within the 24-month study (analysis) period.

![Example of a cumulative AUC measurement from a sample 24-month EDSS/time plot](image)

Figure 1: Example of a cumulative AUC measurement from a sample 24-month EDSS/time plot

The red line represents baseline EDSS. Dots indicate individual EDSS measurements. The area above baseline EDSS (blue) minus the area below baseline EDSS (mauve) equals the cumulative summed AUC (ΔAUC).

This approach to calculating AUC presumes EDSS change between two successive assessments is a relatively sudden, attack-based step or down in EDSS. Comparable studies employing serial disability/time analysis have previously applied a trapezoidal variant of the described approach, which presumes disability to described a constant, linear change between assessment points.\(^1\)\(^-\)\(^3\) The preference for a rectangular approach over a trapezoidal rule is guided by clinical observations that a
sudden attack-based deterioration or regression in EDSS is more probable than one that presumes a constant rate of change. Given the vast majority of the analysed population has relapsing-remitting MS.

However, as a sensitivity analysis, the performance of the rectangular AUC approach was compared with that derived using the trapezoidal rule. The trapezoidal approach employs a two-step approach to calculating AUC change (ΔAUC). First, total area under the curve, including the baseline area, is calculated and then summed using the trapezium rule described in formula 2 to derive the quantity \( \text{AUC}_{\text{sum}} \). \( \Delta \text{AUC} \) is then derived by subtracting the product of the baseline EDSS and the total observation time (24 months) from \( \text{AUC}_{\text{sum}} \) to return the change in baseline AUC (formula 3).

\[
\text{AUC}_{\text{sum}} = \frac{1}{2} \sum (t_{i+1} - t_i)(y_i + y_{i+1}) \tag{2}
\]

\[
\Delta \text{AUC} = \text{AUC}_{\text{sum}} - (y_0)(t_{n} - t_{0}) \tag{3}
\]

Under both approaches, annualised ΔAUC is derived by dividing ΔAUC by two (representing the 2 years of follow-up in this particular analysis) as per formula 4.

\[
\text{Annualised} \ \Delta \text{AUC} = \frac{\Delta \text{AUC}}{2} \tag{4}
\]

This permits comparative analysis of the now annualised AUC change metric by treatment arm to be expressed in terms of “EDSS years”.
C.4.2.1 References


D. ANALYSIS - the timing of relapse onset is seasonal and this relationship is latitude-dependent


Full title
Seasonal variation of relapse rate in multiple sclerosis is latitude-dependent

Running head
Seasonal and latitudinal variation of MS relapses

Authors
Tim Spelman¹, Orla Gray², Maria Trojano³, Thor Petersen⁴, Guillermo Izquierdo⁵, Alessandra Lugaresi⁶, Raymond Hupperts⁷, Roberto Bergamaschi⁸, Pierre Duquette⁹, Pierre Grammond¹⁰, Giorgio Giuliani¹¹, Cavit Boz¹², Freek Verheul¹³, Celia Oreja-Guevara¹⁴, Michael Barnett¹⁵, Francois Grand'Maison¹⁶, Maria Edite Rio¹⁷, Jeannette Lechner-Scott¹⁸, Vincent Van Pesch¹⁹, Ricardo Fernandez Bolanos²⁰, Shlomo Flechter²¹, Leontien Den Braber-Moerland²², Gerardo Iuliano²³, Maria Pia Amato²⁴, Mark Slee²⁵, Edgardo Cristiano²⁶, Maria Laura Saladino²⁷, Mark Paine²⁸, Norbert Vella²⁹, Krisztian Kasa³⁰, Norma Der³¹, Joseph Herbert³², Fraser Moore³³, Tatjana Petkovska-Boskova³⁴, Raed Alroughani³⁵, Aldo Savino³⁶, Cameron Shaw³⁷, Steve Vucic³⁸, Vetere Santiago³⁹, ElizabethAlejandra Bacile⁴⁰, Eli Skromne⁴¹, Dieter Poehlau⁴², Jose Antonio Cabrera-Gomez⁴³, Robyn Lucas⁴⁴ and Helmut Butzkueven¹,⁴⁵.
Affiliations –

1. Department of Neurology, Royal Melbourne Hospital, Parkville, Australia
2. Craigavon Area Hospital, Portadown, Northern Ireland
3. Department of Basic Medical sciences, Neuroscience and Sense Organs, University of Bari, Italy
4. Kommunehospital, Arhus C, Denmark
5. Hospital Universitario Virgen Macarena, Sevilla, Spain
6. MS Center, Department of Neuroscience and Imaging, University ‘G. d’Annunzio’, Chieti, Italy
7. Orbis Medical Centre, Sittard-Geleen, The Netherlands
8. Neurological Institute IRCCS Mondino, Pavia, Italy
9. Hôpital Notre Dame, Montreal, Canada
10. Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada
11. Ospedale di Macerata, Macerata, Italy
12. Karadeniz Technical University, Trabzon, Turkey
13. Groene Hart ziekenhuis, Gouda, The Netherlands
14. University Hospital San Carlos, Madrid, Spain
15. Brain and Mind Research Institute, Sydney, Australia
16. Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
17. Hospital S. Joao, Porto, Portugal
18. John Hunter Hospital, Newcastle, Australia
19. Cliniques Universitaires Saint-Luc, Brussels, Belgium
20. Hospital Universitario Virgen de Valme, Seville, Spain
21. Assaf Harofeh Medical Center, Tzrifin, Israel
22. Francicus Ziekenhuis, Roosendaal, The Netherlands
23. Ospedali Riuniti di Salerno, Salerno, Italy
24. Department of Neurology University of Florence, Florence, Italy
Timothy Denis Spelman (58172)

25 Flinders University and Medical Center, Adelaide, Australia

26 Hospital Italiano, Buenos Aires, Argentina

27 INEBA, Buenos Aires, Argentina

28 St Vincents Hospital, Fitzroy, Australia

29 Mater Dei Hospital, Msida, Malta

30 Jahn Ferenc Teaching Hospital, Budapest, Hungary

31 Hospital Fernandez, Buenos Aires, Argentina

32 NYU Hospital for Joint Diseases, New York, USA

33 Jewish General Hospital, Montreal, Canada

34 JZU Clinic for Neurology, Skopje, Republic of Macedonia

35 Amiri Hospital, Kuwait City, Kuwait

36 Consultorio Privado, Buenos Aires, Argentina

37 Geelong Hospital, Geelong, Australia

38 Westmead Hospital, Sydney, Australia

39 HiGA Gral, San Martin La Plata, Argentina

40 Instituto de Neurociencias Cordoba, Cordoba, Argentina

41 Hospital Angeles Mexico City, Lomas, Mexico

42 Multiple Sclerosis Center Kamillus-Klinik, Asbach, Germany

43 Centro Internacional de Restauracion Neurologica, Havana, Cuba

44 National Center for Epidemiology and Population Health, Australian National University, Canberra, Australia

45 Department of Medicine (RMH), The University of Melbourne, Parkville, Australia

**Corresponding Author**

Tim Spelman;
Abstract

Objective: Previous studies assessing seasonal variation of relapse onset in multiple sclerosis have had conflicting results. Small relapse numbers, differing diagnostic criteria and single region studies limit the generalizability of prior results. The aim of this study was to determine if there is a temporal variation in onset of relapses in both hemispheres and to determine if seasonal peak relapse probability varies with latitude.

Methods: The MSBase international registry was utilized to analyze seasonal relapse onset distribution by hemisphere and latitudinal location. All analyses were weighted for the patient number contributed by each center. A sine regression model was used to model relapse onset and ultra-violet radiation (UVR) seasonality. Linear regression was used to investigate associations of latitude and lag between UVR trough and subsequent relapse peak.

Results: 32,762 relapses from 9811 patients across 30 countries were analyzed. Relapse onset followed an annual cyclical sinusoidal pattern with peaks in early spring and troughs in autumn in both hemispheres. Every 10 degrees of latitude away from the equator was associated with a mean decrease in ultra-violet radiation trough to subsequent relapse peak lag of 28.5 days (95% CI 3.29, 53.71, p=0.028).

Interpretation: We demonstrate for the first time that there is a latitude-dependent relationship between seasonal UVR trough and relapse onset probability peak independent of location-specific UVR levels, with more distal latitude associated with shorter gaps. We confirm prior meta-analyses showing a strong seasonal relapse onset probability variation in the northern hemisphere, and extend this observation to the southern.
Introduction

It is well recognized that the incidence and prevalence of multiple sclerosis (MS) varies significantly with geographical location: it is more common in northern parts of the northern hemisphere and southern parts of the southern hemisphere. The reason for this latitudinal variation is as yet unknown. It has been postulated that climate, particularly sun exposure, could have an impact on the incidence of first demyelinating events and on relapse probability. Vitamin D status (measured as the serum concentration of a metabolite, 25-hydroxyvitamin D, 25(OH)D) is, highly dependent on sunlight exposure, and thus varies according to both latitude and season and has been postulated as the effector in prospective studies.

Several studies have examined seasonal variation of relapse onset probability in multiple sclerosis, with conflicting results. These studies have generally been limited by small numbers of cases and relapses reported. All published studies have reported on single clinical centers. Comparisons of studies of temporal variation of relapse onset are difficult - studies span from 1955 to 2004 and have used differing diagnostic criteria. Inclusion criteria differ with some studies involving cases of definite MS only and others including probable cases. The definition of what constitutes a relapse also differs between studies and both prospective and retrospective studies have been performed.

In 2000 Jin et al performed a meta-analysis of ten studies that met their inclusion criteria: population-based studies including at least thirty cases reporting on season of onset of MS exacerbations. Although some of these individual studies reported no variation in relapse onset, using a weighted mean difference approach, the meta-analysis found a clear seasonal variation of relapse onset, with a spring peak and a winter trough. This temporal fluctuation with spring or summer peak and trough in winter has been subsequently reported in small studies in Spain and Japan, though a further study in the USA failed to confirm a cyclical trend.
The objective of the current study was to re-examine seasonal variation in relapse onset and establish if this temporal variation is influenced by latitude or hemisphere, using the MSBase Registry, a global multi-center cohort study of MS outcomes.

**Materials and Methods**

*The MSBase Registry*

The MSBase Registry is a global, longitudinal, observational registry for Multiple Sclerosis. The registry is a collaborative research group that prospectively collects outcomes data from MS treatment centers using an internet-based, physician owned and operated system [www.msbase.org](http://www.msbase.org). Each center enters patient data in the offline iMed© local electronic database during routine clinic visits and intermittently uploads anonymized datasets to the MSBase server. Physicians record clinical information such as date of MS onset, Kurtzke Expanded Disability Status Score (EDSS), relapse characteristics, MRI and other investigations and diagnostic criteria used. Records are classified as complete and eligible for analyses if they meet a minimum required set of data. Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing center.

*Quality assurance*

The study was performed on datasets from 55 clinical centers. The use of the iMed© electronic database and minimum dataset requirements for inclusion into MSBase ensures a unified approach; cases fulfil the Poser or McDonald criteria for MS and clinical information including relapse data is collected in “real time”. The minimum dataset includes details on patient demography (sex, birth date, MS onset date), patient assessments (visit date, KFS, EDSS), relapses (relapse date, region affected, corticosteroid treatment), para-clinical tests (test date and type including MRI, cerebrospinal fluid examination and biochemistry) and treatment (treatment name, commencement...
and discontinuation dates). The typical relapsing remitting patient attending clinic attends on average every 5.5 months, suggesting that relapse onset documentation is likely to be accurate. Quality assurance was maintained with inbuilt data quality checking in the iMed© local record system, which applied to key dates and data in the minimum dataset. In order to ensure EDSS competency, all participating neurologists completed the Neurostatus certification (http://www.neurostatus.net) or provided evidence of prior completion of this certification.

Inclusion criteria

Centers contributing more than 20 registered patients and who consented to participate in this study were included in the analysis. As at the date of data extract and compilation (1<sup>st</sup> December 2013), a total of 20189 patients from contributing centers were registered in the MSBase registry. Only patients with relapses recorded after commencement of the prospective observation period, defined by the first EDSS score entered into the system, were included in the analysis (n=9811). The follow-up period for each patient was defined as the period commencing at the date of this first recorded EDSS score and ceasing at the date of the last EDSS score entered into the system at the time of data extract. These 9811 patients contributed a total of 32,762 relapses to the analysis. All cases satisfied the Poser criteria for definite MS<sup>24</sup> or the McDonald criteria for MS<sup>25</sup>.

Outcome measures

The two primary outcomes were to determine whether there was temporal variation of relapse onset probability in the northern and/or southern hemispheres and next assess for association between latitude and the time lag between seasonal peak relapse frequency and the preceding seasonal trough in UVR, with the hypothesis that an effect of low vitamin D levels on increased relapse probability would show such an association, as absolute vitamin D levels are lower in regions further away from the equator, and seasonal population vitamin D nadirs are likely reached earlier after the winter solstice.
Relapse dates

The recording of a relapse in the iMed© system requires a complete date to be entered (date, month and year). It is therefore expected that centers would use default dates to record a relapse when the exact date was not known. The relapse data were first analyzed for use of default dates used by centers. Of the 32,762 recorded relapses, 7913 (24.2%) were documented on the first day of a month and 4594 (14.02%) on the 15th day of a month. By comparison the proportion of relapses reported on days of the month other than the 1st and the 15th ranged from 0.8% to 5.6%, representing statistically significantly smaller proportions, suggesting that these days were used as default dates. Relapses recorded on either the 1st or the 15th of the month were therefore randomized to a day within an interval plus or minus 15 days either side of the 1st and the 15th of the month respectively. To assess the validity of this approach we conducted sensitivity analyses to compare the results and performance of the modelling using these randomized dates (as described in the statistical analyses section) against the equivalent modelling for two alternate scenarios – 1) using reported dates without randomization of the 1st and 15th and 2) excluding relapses recorded on the 1st and the 15th of the month altogether. The results of these sensitivity analyses are summarized in the results section. Overall, patients were followed up for a mean (SD) duration of 8.3 years (6.0) until cessation of the study period on the 1st December 2013, the date of data extract and compilation.

Ultraviolet radiation (UVR)

Daily average erythemally weighted ambient UVR for each month for the period 1979-2004 was sourced for all locations included in the analysis.
Statistical analyses

Categorical patient characteristics were summarized using frequency and percentage whilst continuous characteristics were summarized using mean and standard deviation (SD) or median and inter-quartile range (IQR) as appropriate. Patient characteristics were compared between hemispheres using a χ², Wilcoxon rank-sum or t-test as appropriate. Relapses were analyzed separately by hemisphere and latitudinal location. Numbers of relapses per month for each location were calculated. All analyses were weighted for the number of patients contributed by each center. Analysis of plots of observed monthly relapses with confidence intervals at the level of the individual geographic location, hemisphere and across all centers combined suggested a consistent trend towards a spring peak in relapses (March-April in the northern hemisphere, September-October in the southern hemisphere) with an autumn trough (September-October in the northern hemisphere, March-April in the southern). This suggested a regression model consisting of one sine and one cosine function, describing a single annual cycle with one peak and one trough separated by six months, as most appropriate for capturing the observed seasonality in relapse onset. We supplemented visual plot analysis with an Edwards’ test to test the distribution of monthly relapses for significant departures from a uniform distribution. We formally tested the assumption of an underlying single annual cycle described by a single sine and single cosine function by extending the geometric regression models to include additional sine and cosine terms to test for sub-annual cycles, at both the level of the hemisphere and the individual location. Models incorporating additional period harmonics were compared against the base model of a single sine and cosine function using analysis of residual plots, a likelihood ratio test and adjusted coefficient of determination (R²). The base model across northern hemispheric locations minimized the residual square error and returned a superior fit (p<0.0001, adjusted R² = 0.263) when compared to either a model incorporating an additional harmonic ((p=0.0001, adjusted R² = 0.198) or an additional two harmonics (p=0.0014, adjusted R² = 0.181). The same base model applied across locations in the southern hemisphere returned a comparable fit (p<0.0001, adjusted R² = 0.241) with a model
incorporating an additional harmonic \((p<0.0001, \text{ adjusted } R^2 = 0.243)\), with the extended model only marginally reducing the residual square error when compared with the base. A further extension of the southern hemisphere model incorporating an additional two harmonics returned an inferior fit of the observed data \((p<0.0001, \text{ adjusted } R^2 = 0.167)\) when compared to the base model. As a final check of the appropriateness of the base model for southern hemisphere locations, we re-estimated phase shift and peak relapse probability date using the alternate base plus one additional harmonic model as a further sensitivity analysis. This extended model returned exactly the same point estimate of southern hemispheric relapse peak \((T_{max})\) as the base model \((5^{th} \text{ September})\) with only a marginal contraction of the associated 95% confidence interval \((10^{th} \text{ August} – 25^{th} \text{ September} \text{ in the extended model compared with } 10^{th} \text{ August} – 26^{th} \text{ September in the base model})\). At the level of the individual geographic location, the base model consistently outperformed either of the extended harmonic models for all individual locations, both in terms of minimizing the residual square error and maximizing \(R^2\). This process was repeated for the UVR data, which similarly was best described using a single annual cycle with a single peak and trough separated by six months. This process was repeated for the UVR data, which similarly was best described using a single annual cycle with a single peak and trough separated by six months.

As monthly relapses were significantly skewed, this outcome was log-transformed. The log-transformed quantity was itself assessed for skew using a Shapiro-Wilk test. Transformed monthly relapses were approximately normally distributed, permitting parametric sine-regression. We tested for interaction between sine and cosine terms and hemisphere. Location-specific median monthly UVR was incorporated as a fixed effect into the sine regression model to adjust our estimates of peak relapse probability date for absolute UVR levels at each site. Phase-shift was estimated using double-angle formulae. The phase-shift is the estimate of the point on the sine curve at which the curve crosses the mid-point and starts to climb towards its peak three months later. Thus estimation of the phase shift permitted calculation of the estimated date of both the peak \((T_{max})\) and
subsequent trough (Tmin) in monthly relapses or UVR as appropriate (where Tmax = phase-shift + (365/4) and Tmin = phase-shift + ((365/4)*3)). To explore the distribution, consistency and magnitude of seasonality between sites and latitudes we used the sine regression model to calculate the difference between the estimated number of relapse per 100 patients at Tmax (peak relapse probability) and Tmin (trough relapse probability) for each site. Peak-to-trough difference was then compared between sites and ranges of absolute latitude.

Using the sine regression models to estimate peak and trough dates for both relapses and UVR enabled calculation of the time lag from UVR trough to subsequent relapse peak. The distribution of the relationship between lag and latitude was tested for both linearity and non-linearity with a linear relationship demonstrating the best fit with the smallest associated error. The UVR trough-to-relapse peak lag was tested for significant skew using the Shapiro-Wilk test. As lag was approximately normally distributed, linear regression of the mean was used to test for association between location latitude and lag, weighting for the number of patients at each clinic location. Goodness of fit was assessed through analysis of the coefficient of determination.

As a further sensitivity analysis we tested for the influence of patient-level propensities for relapse by rerunning all modelling using a mixed effects Poisson regression derived on a single sine and cosine function and incorporating patient as a random effect. Fixed effects included in the model were hemisphere, location and absolute location-specific median monthly UVR in addition to potential patient-level predictors of relapse including age at disease onset, baseline EDSS and prior exposure to DMD treatment (quantified as the proportion of follow-up time pre-relapse treated with DMD). Double-angle formulae were again used to calculate phase shift which in turn enabled estimation of peak relapse probability date and, as a result, UVR trough to relapse peak lag.
All reported p values are exact and 2-tailed, and for each analysis p<0.05 was considered significant. All analyses were conducted in Stata version 12.0 (StataCorp, College Station, Texas) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Data were extracted on 1st December 2013 from the MSBase registry, which resulted in the inclusion of data from 46 clinical centers in 20 countries. The 9811 included patients contributed 32,762 documented post-first EDSS relapses. The female to male ratio was 2.6:1. Median age at disease onset was 28.9 years (inter-quartile range: 23.1, 36.3). 7732 (78.8%) of patients had relapsing remitting MS as at the date of extract, the remainder had secondary progressive MS. Of the 46 clinical centers included in the analysis, 31 in the northern hemisphere contributed 29,142 (89.0%) relapses from 8411 patients and the remaining 15 clinics in the southern hemisphere contributed 3620 (11%) relapses from 1400 individual patients. The median (IQR) number of relapses observed per patient across the follow-up period was 2 (1, 4) whilst the mean (SD) proportion of follow-up time on treatment per patient was 0.6 (0.4). Overall, mean (SD) follow-up per patient was 8.3 years (6.0). There were no significant differences in any of these characteristics by hemisphere (Table 1).

<table>
<thead>
<tr>
<th>Patient characteristic</th>
<th>Level</th>
<th>Total (n=9811)</th>
<th>Northern hemisphere (n=8411)</th>
<th>Southern hemisphere (n=1400)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex – n (%)</strong></td>
<td>Female</td>
<td>7076 (72.1)</td>
<td>6042 (71.8)</td>
<td>1034 (73.9)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>2735 (27.9)</td>
<td>2369 (28.2)</td>
<td>366 (26.1)</td>
</tr>
<tr>
<td><strong>Age at disease onset (years) - median (IQR)</strong></td>
<td>-</td>
<td>28.9 (23.1, 36.3)</td>
<td>28.8 (23.1, 36.1)</td>
<td>29.2 (23.6, 36.5)</td>
</tr>
<tr>
<td><strong>Baseline EDSS - median (IQR)</strong></td>
<td>-</td>
<td>2 (1.5, 3.5)</td>
<td>2 (1.5, 3.5)</td>
<td>2.5 (1.5, 3.5)</td>
</tr>
<tr>
<td><strong>Number of relapses observed per patient during follow-up - median (IQR)</strong></td>
<td>-</td>
<td>2 (1, 4)</td>
<td>2 (1, 4)</td>
<td>2 (1, 4)</td>
</tr>
<tr>
<td><strong>Proportion of follow-up time on treatment - mean (SD)</strong></td>
<td>-</td>
<td>0.6 (0.4)</td>
<td>0.6 (0.4)</td>
<td>0.6 (0.4)</td>
</tr>
<tr>
<td><strong>Follow-up duration (years) - mean (SD)</strong></td>
<td>-</td>
<td>8.3 (6.0)</td>
<td>8.3 (6.0)</td>
<td>8.1 (6.1)</td>
</tr>
</tbody>
</table>

Table 1: Patient characteristics
a. The distribution of patient characteristics by hemisphere were compared using a χ² test, Wilcoxon rank-sum or t-test as appropriate. All comparisons were associated with a p>0.05.
“Baseline” refers to the commencement of the prospective observation period for each patient, defined by the first EDSS score entered into the system. The follow-up period for each patient refers to the time lapsed from baseline to the last recorded EDSS score entered into the system.

**Temporal variation**

Relapse onsets followed an annual cyclical pattern and were most common in spring and least common in autumn in both hemispheres. There was no difference in this pattern by hemisphere (test of interaction $p=0.254$). This seasonal variation was statistically significant, with the phase-shift across all relapses estimated at $-24.83$ (95% CI $-45.81$, $-3.92$) which translated into an estimated relapse peak (Tmax) of the 7th March (95% CI 10th February, 28th March) in the northern hemisphere and 5th September (95% CI 10th August, 26th September) in the southern hemisphere. The sine-wave regression of relapse onset in both northern and southern hemispheres are shown in Figure 1 and a spider-web plot of relapses by month combined across both hemispheres is shown in Figure 2. Across all sites and latitude, mean (SD) peak-to-trough relapse difference was 7.62 (6.57) relapses per 100 patients. Across both hemispheres, centers at latitudes of 40 or more degrees, larger differences in peak-to-trough difference (mean 8.62 relapses per 100 patients, SD 7.64) were observed compared with sites located within an absolute latitude range of 20 to 39 degrees (mean 5.71 relapses per 100 patients, SD 3.30). However, this difference was not significant ($p=0.135$).
Figure 1: Sine-wave regression of number relapse onsets by month in the northern and southern hemispheres.
Figure 2: Spider-web plot of distribution of monthly relapses combined across both hemispheres. Month labels ordered as northern/southern hemisphere.

**Latitudinal variation and UVR**

The lag between UVR trough and relapse onset peak varied with study center latitude. Mean (SD) lag was 2.72 months (1.61). Increasing latitude (away from the equator) was associated with a statistically significant decrease in the time between the UVR trough and the subsequent relapse peak. Every 10 degree increase of latitude away from the equator was statistically significantly associated with a mean decrease in UVR trough to relapse peak lag of 28.5 days (95% CI 3.29, 53.71, p=0.028), weighting for the number of patients per location and controlling for absolute UVR at each location (Fig 3). There was no difference in this association by hemisphere (p=0.81).

A sensitivity analysis using the original reported dates rather than randomizing relapse onsets recorded on the 1st or 15th of the month as described above yielded similar results. Peak relapse onset in the northern hemisphere was estimated to occur 3 days earlier on the 4th March (95% CI February 8th, March 31st), whilst every 10 degree increase in latitude away from the equator remained predictive of a similar magnitude of decrease in lag between UVR trough and subsequent relapse peak (27.44 day decrease in lag, 95% CI 1, 54, p=0.042). Rerunning the modelling excluding all relapses reported on either the 1st or 15th of any month resulted in a marginally later relapse peak (March 12th in the northern hemisphere) and a significant, albeit smaller, decrease in lag with increasing latitude (21.90 day decrease in lag, 95% CI 0.21, 43.60, p=0.048).

As a further sensitivity analysis, we supplemented these location-level models with a patient-level mixed effects Poisson model to explore the potential influence of patient-level propensities towards relapse on our observed seasonal and latitudinal trends. Consistent with the location-level analysis, seasonal variation was again statistically significant under the patient-level model, with phase-shift
across all relapses estimated at -22.91 (95% CI -44.18, -2.01). This translated into an estimated relapse peak ($T_{\text{max}}$) of the 9th March (95% CI 9th February, 29th March) in the northern hemisphere and the 7th September (95% CI 11th August, 27th September) in the southern hemisphere, just two days later than that estimated under the primary analysis model. There was no difference by hemisphere (test of interaction $p=0.303$). Similarly more distal latitude was again associated with a statistically significant decrease in the time from UVR trough to subsequent relapse peak. Specifically, every 10 degrees of latitude away from the equator was statistically significantly associated with a mean decrease in UVR trough to relapse peak of 24.8 days (95% CI 1.97, 49.22, $p=0.032$). This represents a modest reduction in this lag of a mean 4.1 days when compared with the primary analysis. There was no difference in this association by hemisphere (test of interaction $p=0.671$).

Figure 3: Weighted line of best fit between latitude and UV trough-to-relapse peak with observed coordinates representing individual centres scaled to reflect relative number of patients.
Discussion

In our large, multinational study of MS outcomes, we confirm and extend a prior meta-analysis by showing a strongly seasonal relapse probability in both hemispheres. We demonstrate for the first time that there is a latitude-dependent lag between seasonal UVR trough and relapse probability peak, with increasing latitudes away from the equator associated with shorter lag. Our lead hypothesis to explain this phenomenon is that people with MS living at latitudes further away from the equator, who are known to have significantly lower vitamin D status (largely UVR determined) at all seasons, reach a “low” threshold level of 25(OH)D or of another UVR-mediated, long half-life immunomodulatory effect sooner after the winter solstice than people residing at more equatorial latitudes.

Many previous studies of seasonal variation in relapse onset have been difficult to interpret; they were performed over a 60-year period, have small case numbers and differing methodology. On an individual basis, they seem to provide conflicting results. However, results of the meta-analysis performed by Jin in 2000⁹ are confirmed and extended by our multi-center multinational analysis, showing an overall peak in the probability of relapse onset in spring and trough in winter. The large overall cohort size lends such analyses power to detect significant effects, as does the prospective assessment that avoids recall bias.

Our observation that contributing centers located at a more distal latitude away from the equator significantly predicts a shorter lag between seasonal trough UVR levels in mid-winter and subsequent peak in relapse probability was consistent across hemispheres. Several season-related factors could influence relapse probability in this way, including those commonly discussed such as vitamin D status (influenced by ambient UVR), non-vitamin D immunomodulatory effects of UVR, and viral illnesses, particularly upper respiratory infections. We believe that among these, a seasonal
variation of lag between winter UVR trough and relapse peak is best explained by an effect of changing vitamin D status and/or other UVR-related direct immune influence with a several-week duration to clinical symptomatology. The lowest troughs for both of these measures would be expected in early spring and would occur significantly earlier in latitudes with the lowest autumn and winter sun and UVR dose. Of course, UVR-mediated non vitamin-D immunological effects of long half-life would fit the same pattern and could in themselves be very indirect, e.g. by increasing the probability of seasonal viral infections slightly.

In support of our hypothesis, a number of recent studies have suggested that low vitamin D status is widespread throughout mid- to high-latitude locations with a late winter/early spring trough. Shoben et al showed, using data from a cardiovascular health study conducted in 4 locations in the USA and acquired in 1992/1993, that vitamin D status was cyclical, with trough levels between February and April. In a prospective population-based cohort study of MS patients, the large majority of whom were on immunomodulatory therapy, higher vitamin D status was associated with a lower probability of relapse, even after adjusting for seasonality. An inverse relationship between relapse incidence and vitamin D status has also been identified in both adult and pediatric onset MS patients. Consistent with these results, Auer et al demonstrated annual sinusoidal variation in the number of gadolinium-enhancing MRI lesions, proposing that fluctuations in sunlight and UVR exposure, in addition to the seasonal variation in common viral infections, could explain the seasonality observed for MRI lesion activity. Mowry et al also demonstrated an inverse relationship between vitamin D status in individuals and disease activity on brain MRI, with each 10ng/ml (25nmol/L) increase in 25(OH)D levels associated with a 15% lower risk of a new T2 lesion and 32% lower risk of a gadolinium-enhancing lesion. In another study, increased summer sun exposure was associated with increased grey matter and whole brain volume, independent of 25(OH)D levels. The above studies support suggestions of both vitamin-D dependent and independent seasonal effects potentially acting independently to influence relapse and MRI lesion probability.
Other environmental factors with seasonal and latitudinal gradients, such as clinical infections, could also influence relapse probability. Upper respiratory tract infections (URTIs) such as influenza and the common cold describe seasonal fluctuations with peak incidence at latitudes further away from equator\textsuperscript{38}. Compared with our modelled observation of an early spring peak in relapse, the incidence of URTIs exhibit largely winter peaks\textsuperscript{38}. However infections such as measles and varicella typically record a similar peak incidence in late-winter to early spring, particularly in temperate climates\textsuperscript{39,40}. Complicating this, maintenance of serum 25(OH)D levels above 38 ng/ml has been previously observed in a prospective cohort study (n=195) to decrease the risk of acute respiratory tract infection in healthy non-MS patients (risk ratio 0.51, p<0.0001)\textsuperscript{41}, so that indirect effects are also possible. In MS patients, clinical infections have previously been hypothesized to increase the probability of relapse via increased release of inflammatory mediators\textsuperscript{42-46}. A relatively small (n=73), but prospective, longitudinal study of the association between infection and relapse in RRMS patient observed an increased risk of relapse (rate ratio 2.1) within the time period proximal to an infection (two weeks prior to 5 weeks post a clinically manifest infection), where the majority of infection events were URTIs\textsuperscript{47}. Whilst our own study did not collect data pertaining to clinical infections and their potential influence on relapse probability, these studies underscore the importance of large, clinical trials that can control and adjust for infection events to better isolate the change, if any, in relapse probability attributable to 25(OH)D.

In addition to potential infectious factors, several immunomodulatory influences of relapse probability, both 25(OH)D and non-25(OH)D-mediated, have been hypothesized\textsuperscript{48}. 25(OH)D has been demonstrated to shift T helper (Th) -lymphocytes from a pro-inflammatory Th1 profile to a Th2 profile associated with reduced IFNγ\textsuperscript{49}, IL-1\textsuperscript{49,50}, IL-2\textsuperscript{49,51} and IL12\textsuperscript{49,50} levels whilst inducing the IL-10 production\textsuperscript{49,52}. 25(OH)D has further been observed to inhibit T-helper type 1 and dendritic cells whilst inhibiting IgM and IgG antibody production from plasma cells\textsuperscript{51,53-56}. Protective effects of
25(OH)D have additionally been observed in experimental autoimmune encephalitis, a commonly used animal model of MS\textsuperscript{57-59}. Amongst non-25(OH)D immunomodulatory effects, of particular recent interest has been the observation that UVR-irradiation of skin alters dendritic cell progenitors in the bone marrow via epigenetic mechanisms mediated through prostaglandin E2. Reduction in antigen-presentation competency is sustained in the dendritic cells derived from these progenitor cells\textsuperscript{60}. Evidence for this pathway has been extensively reviewed previously\textsuperscript{61-63}.

The major limitation of studies such as ours is that they are only hypothesis-generating in this regard, as individual serial assessments including serum vitamin D tests, skin UVR exposure assessments and a meticulous prospective record of viral infections are beyond the scope of our global outcomes registry. Whilst many of the studies described herein correlate 25(OH)D levels with an inverse probability of relapse, some adjusting for seasonality, they do not however adjust or control for non-25(OH)D, UVR-mediated effects. The best way to prove or disprove, or otherwise isolate, a hypothesized causal relationship between 25(OH)D and relapse probability is via clinical trials. Two of three recent trials have been published with encouraging results, although all are likely to be under-powered. A Finnish add-on trial of 25(OH)D to IFNβ-1b observed a significant reduction in the frequency of new T1 gadolinium-enhancing cerebral lesions in the 25(OH)D treatment arm (n=33/arm)\textsuperscript{64}. A second Iranian pilot trial randomized serum 25(OH)D-deficient (<30 ng/ml) CIS patients with optic neuritis to either 50,000 IU of Vitamin D3 or placebo weekly for 12 months (n=15/arm). Relative risk of relapse in the treatment arm was reduced by 68.4% (relative risk=0.326, p=0.007) whilst a trending benefit in MRI measures of disease activity was observed in the D3 treatment group\textsuperscript{65}. On the other hand, a small Australian randomized trial of low-versus high dose D2 supplementation, largely as add–on treatment to injectable disease-modifying drugs (n=10 or 11/arm), failed to confirm a difference in relapse rate of MRI activity\textsuperscript{66}.
Of note, two large 25(OH)D monotherapy trials are currently underway. The PREVANZ trial (registration ACTRN12612001160820) is an Australian and New Zealand trial randomizing CIS patients to either 1000, 5000 or 10,000 IU/day of 25(OH)D or placebo. The second is a French randomized controlled trial ("D-lay MS", registration PHRC-N/2012/ET) comparing placebo against 100,000 IU/fortnight of 25(OH)D. In addition a further randomized, add-on trial of 25(OH)D or placebo on to sub-cutaneous IFNβ-1a (SOLAR) is in progress. If these large clinical trials corroborate the theorized inverse relationship between 25(OH)D and relapse probability, the dosing schedules used may provide a basis for guiding a location-appropriate 25(OH)D supplementation therapy in terms of quantity, frequency and timing, given the suggestions of seasonal and latitudinal trends observed in our study.

**Acknowledgements**

The authors would like to thank Ivan Hanigan for his support in extracting and interpreting the ultraviolet radiation satellite data. The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. We wish to dedicate this paper to the memory of Professor Damien Jolley, an inspirational biostatistician and epidemiologist who assisted with the original modeling and died on 15/01/2012.

**Disclosures**

Tim Spelman received honoraria for consultancy and funding for travel from Biogen Idec Inc, Orla Gray did not declare any competing interests, Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva, Thor Petersen received funding or
speaker honoraria from Biogen Idec, Merck Serono, Novartis, Bayer Schering, Sanofi-Aventis, Roche, and Genzyme, Guillermo Izquierdo received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva, Alessandra Lugaresi is a Bayer Schering, Biogen Idec, Genzyme/Sanofi, Merck Serono Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi and Teva and research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva. Prof Lugaresi has also received travel and research grants from the Associazione Italiana Sclerosi Multipla and was a Consultant of “Fondazione Cesare Serono”, Raymond Hupperts received honoraria as consultant on scientific advisory boards from Merck-Serono, Biogen-Idec, Genzyme-Sanofi and Teva, research funding from Merck-Serono and Biogen-Idec, and speaker honoraria from Sanofi-Genzyme, Roberto Bergamaschi, Pierre Duquette has received honoraria for organising CME events and has obtained funding to attend meetings from Biogen Idec, EMD Serono, TEVA Neuroscience, Novartis, and Genzyme, has received funding for investigator-initiated trials with Biogen Idec, EMD Serono and Novartis, and has received peer-review funding from CIHR and from the MS Society of Canada, Pierre Grammond is a Novartis, Teva-neuroscience, Biogen Idec advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis, Giorgio Giuliani did not declare any competing interests, Cavit Boz has received travel grants from Merck Serono, Biogen Idec, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis., Freek Verheul is an advisory board member for Teva Biogen Merck Serono and Novartis, Celia Oreja-Guevara received honoraria as consultant on scientific advisory boards from Biogen-Idec, Bayer-Schering, Merck-Serono, Teva and Novartis; has participated in clinical trials/other research projects by Biogen-Idec, GSK, Teva and Novartis, Michael Barnett has received honoraria for participation in advisory boards and travel sponsorship from Novartis, BioCSL, Genzyme and Biogen Idec, Francois Grand’Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from
Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014, Maria Edite Rio did not declare any competing interests, Jeannette Lechner-Scott has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen Idec, CSL, Genzyme Sanofi, Merck Serono and Novartis., Vincent Van Pesch has served on advisory boards for Biogen Idec and Genzyme; has received travel grants from Biogen Idec, Bayer Schering, Sanofi Aventis, Merck Serono and Novartis Pharma; has received consultancy fees from Biogen Idec, Teva and Novartis Pharma; has received research grants from Bayer Schering, Ricardo Fernandez Bolanos did not declare any competing interests, Shlomo Flechter did not declare any competing interests, Leontien Den Braber-Moerland did not declare any competing interests, Gerardo Iuliano had travel/accommodations/meeting expenses funded by Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi Aventis, and Teva, Maria Pia Amato received honoraria as consultant on scientific advisory boards by Biogen-Idec, Bayer-Schering, Merck-Serono, Teva and Sanofi-Aventis; has received research grants by Biogen-Idec, Bayer-Schering, Merck-Serono, Teva and Novartis, Mark Slee has participated in, but not received honoraria for, advisory board activity for Biogen Idec, MerckSerono, BayerSchering, Sanofi Aventis and Novartis, Edgardo Cristiano received honoraria as consultant on scientific advisory boards by Biogen-Idec, Bayer-Schering, Merck-Serono, Genzyme and Novartis; has participated in clinical trials/other research projects by Merck-Serono, Roche and Novartis, Maria Laura Saladino did not declare any competing interests, Mark Paine did not declare any competing interests, Norbert Vella received compensation for travel and honoraria from Novartis, Biogen Idec, Glaxo-Smith-Kline, Krisztian Kasa did not declare any competing interests, Norma Deri did not declare any competing interests, Joseph Herbert did not declare any competing interests, Fraser Moore has participated in clinical trials sponsored by EMD Serono and Novartis, Tatjana Petkovska-Boskova did not declare any competing interests, Raed Alroughani received honororia from Biologix, Bayer, Merck Sorono, GSK and Novartis, and served on advisory board for Biologix, Novartis and Merck Sorono, Aldo Savino did not declare any competing interests, Cameron
Shaw did not declare any competing interests, Steve Vucic did not declare any competing interests, Vetere Santiago did not declare any competing interests, Elizabeth Alejandra Bacile did not declare any competing interests, Eli Skromne did not declare any competing interests, Dieter Poehlau did not declare any competing interests, Jose Antonio Cabrera-Gomez did not declare any competing interests, Robyn Lucas did not declare any competing interests and Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.

References


E. Analysis: Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis

E.1 Quantifying risk of early relapse in patients with first demyelinating events (submitted paper)

TITLE: Quantifying risk of early relapse in patients with first demyelinating events

AUTHORS: Tim Spelman¹ (MBBS), Claire Meyniel¹-² (MD), Juan Ignacio Rojas³ (MD), Alessandra Lugaresi⁴ (MD), Guillermo Izquierdo⁵ (PhD, MD), Francois Grand’Maison⁶ (MD), Cavit Boz⁷ (MD), Raed Alroughani⁸ (PHD, MD), Eva Havrdova⁹ (PhD, MD), Dana Horakova⁹ (PhD, MD), Gerardo Iuliano¹⁰ (MD), Pierre Duquette¹¹ (MD), Murat Terzi¹² (MD), Pierre Grammond¹³ (MD), Jose Antonio Cabrera-Gomez¹⁴ (MD), Raymond Hupperts¹⁵ (MD), Jeannette Lechner-Scott¹⁶ (MD), Celia Oreja-Guevara¹⁷ (PhD, MD), Ricardo Fernández Bolaños¹⁸ (MD), Eugenio Pucci¹⁹ (PhD, MD), Freek Verheul²⁰ (MD), Marcela Fiol²¹ (MD), Vincent Van Pesch²² (PhD, MD), Edgardo Cristiano³ (MD), Thor Petersen²³ (MD), Tatjana Petkovska-Boskova²⁴ (MD), Fraser Moore²⁵ (MD), Ilya Kister²⁶ (MD), Tomas Kalincik¹ (PhD, MD), Vilija Jokubaitis¹ (PhD), Maria Trojano²⁷ (MD); Helmut Butzkueven¹²,²⁸ (PhD, MD), on behalf of the MSBasis (an MSBase substudy) Investigators

¹ Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia

² Department of Neurophysiologie, Hopital Pitié-Salpêtrière, Paris, France

³ Hospital Italiano, Buenos Aires, Argentina

⁴ MS Center, Department of Neuroscience and Imaging, University ‘G. d’Annunzio’, Chieti, Italy

⁵ Hospital Universitario Virgen Macarena, Sevilla, Spain

⁶ Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada

⁷ Karadeniz Technical University, Trabzon, Turkey
Timothy Denis Spelman (S8172)

8 Amiri Hospital, Kuwait City, Kuwait

9 Department of Neurology and Center of Clinical Neuroscience, 1st Faculty of Medicine, General University Hospital and Charles University in Prague, Czech Republic

10 Ospedali Riuniti di Salerno, Salerno, Italy

11 Hôpital Notre Dame, Montreal, Canada

12 Mayis University, Samsun, Turkey

13 Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada

14 Centro Internacional de Restauracion Neurologica, Havana, Cuba

15 Maaslandziekenhuis, Sittard, The Netherlands

16 John Hunter Hospital, Newcastle, Australia

17 Hospital Clínico San Carlos, Madrid, Spain

18 Hospital Universitario Virgen de Valme, Sevilla, Spain

19 Neurology Unit, ASUR Marche – AV3, Macerata, Italy

20 Groene Hart Ziekenhuis, Gouda, the Netherlands

21 Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia, Buenos Aires, Argentina

22 Cliniques Universitaires Saint-Luc, Brussels, Belgium

23 Kommunehospitalet, Aarhus, Denmark

24 JZU Clinic for Neurology, Skopje, Republic of Macedonia

25 Jewish General Hospital, Montreal, Canada

26 New York University School of Medicine, New York, United States

27 Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy

28 Department of Neurology, Box Hill Hospital, Monash University

*Authors contributed equally
Corresponding Author: Tim Spelman

Postal Address:
Melbourne Brain Centre, Royal Melbourne Hospital,
Grattan St, Parkville, Victoria, Australia 3050
Ph: +61 3 9342 8070
Fax: +61 3 9342 8070
Email: tim@burnet.edu.au

Abstract word count: 250
Main body word count: 3156
Number of references: 30
Keywords: clinically isolated syndrome, clinically definite multiple sclerosis, second attack, disease-modifying drugs, MRI, CSF, MS, CIS, nomogram, calibration.
Abstract

Background: Clinical relapse after a first demyelinating event defines the conversion from clinically isolated syndrome (CIS) to clinically definite multiple sclerosis (MS). Not all patients with CIS convert to MS. Studies quantifying time to conversion in CIS patients have informed MS diagnostic criteria. Characteristics at CIS examination assist in identification of patient subsets who are at highest risk of early second attack, and could benefit the most from early disease-modifying drug (DMD) treatment. We prospectively examined determinants of second attack risk in a cohort of 3296 CIS patients.

Methods: Patients with CIS were prospectively followed in the MSBase Incident Study (MSBASIS). Predictors of time to conversion were analysed using Cox proportional hazards regression. Prognostic nomograms were then derived to permit individualised assessment of conversion risk at the time of CIS.

Findings: 3296 patients from 50 clinics in 22 countries were followed up for a median (IQR) of 1.92 years (0.90, 3.71). A total of 1953 (59.3%) patients recorded a second attack. Higher EDSS, brainstem or supratentorial first symptom location, the presence of oligoclonal bands, 1+ T1 gadolinium enhancing lesions, 3+ periventricular lesions, 1+ infratentorial and 1+ juxtacortical lesion on baseline MRI were all predictors of conversion. Conversely, older age at CIS and exposure to DMD post-CIS were associated with reduced rates.

Interpretation: This multinational study shows that age at CIS onset, DMD exposure, EDSS, multiple baseline cerebral MRI criteria and the presence of oligoclonal bands on CSF examination are all independently associated with shorter time to relapse after CIS.
Introduction:
The first relapse after clinically isolated syndrome (CIS) marks the conversion from CIS to clinically definite multiple sclerosis (CDMS). Conversion is common, with published estimates ranging from 20% in those with normal onset cerebral MRI to 75-88% in patients with a high number of MRI lesions combined with various lesion location criteria.\textsuperscript{1-6} Establishing optimal timing for initiating disease-modifying drug (DMD) therapy following CIS is difficult.\textsuperscript{7} Early identification of patients at high risk of conversion can assist this treatment decision,\textsuperscript{8} as therapy is often recommended for high risk CIS patients.\textsuperscript{7,9} Characteristics at the time of CIS including age,\textsuperscript{10} lesion number and distribution characteristics on cerebral MRI\textsuperscript{1,3} and the cerebrospinal fluid (CSF) examination assist in identification of patients at higher risk of early second attack, and thus those patients who may benefit from early treatment.\textsuperscript{11} The DMD’s IFNβ-1a, IFNβ-1b and glatiramer acetate have been trialled in high-risk CIS patients compared to placebo, and all have been shown to significantly reduce the proportion of patients converting to CDMS.\textsuperscript{12-15}

MRI examination and CSF analysis are recommended investigations of CIS as they increase the specificity of diagnosis.\textsuperscript{10,16} Baseline cerebral MRI parameters are well established as predictors of conversion to CDMS.\textsuperscript{1,3,17} The presence of OCB on CSF examination supports a diagnosis of MS\textsuperscript{13} and has previously been demonstrated to increase the risk of conversion in patients with CIS, independent of cerebral MRI parameters,\textsuperscript{18,19} and assists the clinician in differentiating multiple sclerosis from alternate diagnoses, such as infection or vasculitis.\textsuperscript{20,21} A systematic review and meta-analysis of 48 studies concluded that the presence of OCB in CSF in CIS patients was a strong predictor of time to CDMS.\textsuperscript{22} In further confirmatory work, a recent prospective cohort study exploring a comparable suite of demographic, clinical and examination factors at CIS to our study identified MRI lesion number and the presence of CSF oligoclonal bands as high- and medium-impact prognostic factors, respectively.\textsuperscript{23}
The objective of this study was to utilise the prospectively documented CIS cohort from 50 MSBase centres to first examine demographic, clinical, diagnostic and treatment characteristics as predictors of time to second attack following CIS and then to use these data to create and internally validate a predictive nomogram to calculate individualized risk of conversion to CDMS at 12 months.

MATERIALS & METHODS

**MSBase Registry**

The MSBase Registry is an international online database accumulator that was established in 2004 and collects disease related information from consenting patients attending Multiple Sclerosis (MS) clinics. The registry is a collaborative research group that prospectively collects a defined minimum dataset from MS patient treated at specialised MS centres, using an internet-based, physician owned and operated system [www.msbase.org](http://www.msbase.org). The MSBase registry was approved by the Melbourne Health Human Research Ethics Committee, and by the local ethics committees in all participating centres (or exemptions granted, according to applicable local laws and regulations). Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing centre.

**MSBASIS**

The MSBase Incident Study (MSBASIS) is a sub-study of the MSBase registry. Commencing on the 18th November 2004, MSBASIS is an ongoing global, longitudinal, investigator-initiated and maintained observational cohort study designed to prospectively assess all registry MS patients with a clinically isolated syndrome with symptom onset less than 12 months from the enrolment date. As of the 2nd April 2014, the date of data extract and compilation, the study had enrolled 4313 patients from 50 clinics across 22 countries contributing 37,569 clinic visit observation points.
Inclusions

MSBASIS requires a minimum entry dataset at each clinic visit observation point consisting of the visit date, Neurostatus EDSS, Kurtzke functional systems score (KFS), onset date of prospectively observed relapses, glucocorticoid therapy for relapses and, where applicable, initiation and discontinuation dates for MS disease-modifying therapy. EDSS recorded within 30 days of CIS onset were excluded. The first cerebral magnetic resonance imaging (MRI) scan classification for each patient, using Barkhof-Tintore criteria for lesion dissemination in space\textsuperscript{1,3,25,26}, which had to be performed within 12 months of CIS onset, was required. Following the initial baseline visit, minimum annual follow-up was required. Patients with primary progressive MS (PPMS) were excluded. Baseline was defined as the recorded date of CIS onset. Of the 4313 patients enrolled in MSBASIS, a total of 3296 patients satisfied all inclusion criteria and were thus eligible for this study.

Outcome measure & definitions

The primary outcome of this analysis was time to first relapse following clinically isolated syndrome (CIS). Clinically definite MS (CDMS) was defined as examination evidence of a symptomatic second neurological episode attributable to demyelination of more than 24 hours duration and more than 4 weeks from the initial attack, according to the Poser criteria\textsuperscript{27}. Follow up time was defined as the time that lapsed between the date of CIS onset (baseline) and either the date of first post-CIS relapse or, where no subsequent post-CIS relapse was observed, the date of the last recorded clinic visit. First DMDs initiated during follow-up included in the analysis were IM-IFNβ-1a, SC-IFNβ-1a, IFNβ-1b, glatiramer acetate, natalizumab and fingolimod. Medication possession ratio (MPR) was defined as the number of follow-up days on DMD treatment divided by the total number of follow-up days contributed by a patient and was calculated at the level of the individual patient. Baseline CSF examination was included if a sample was collected and tested within 6 months of the baseline date. Lesion number on baseline spinal MRI, if performed, were analysed as both a continuous
variable (number of T1 gadolinium enhancing and T2 hyperintensive lesions) and a categorical variable (zero compared with 1+ lesions).

**Statistical analyses**

Sex, country, age at CIS onset, DMD exposure, medication possession ratio, identity of first DMD product initiated on follow-up, baseline MRI (both cerebral and spinal) and the presence of OCB in baseline CSF examination were analysed for association with time to first relapse following CIS. Whilst a unified, standardised protocol is used by all participating centres for date entry into the registry and the requirement for neurostatus certification ensure a degree of consistency in data capture and outcome ascertainment, all models presented were further adjusted for country to better control for any residual inter-country heterogeneity. Both baseline spinal MRI and OCB status were not required for inclusion. Categorical variables were summarised using frequency and percentage. Continuous variables were assessed for significant departure from normality using a Shapiro-Wilk test and summarised using mean and standard deviation (SD) or median and inter-quartile range (IQR), as appropriate. Kaplan-Meier survival curves were used to describe cumulative time without relapse over the observation period. Cox proportional hazards regression was used to investigate correlation between our a priori identified predictors and time to first post-CIS relapse. Interactions between predictor variables in the adjusted model were tested. Hazard proportionality was assessed through analysis of scaled Schoenfeld residuals. The linearity of association between candidate explanatory variables and the conversion end-point were tested by incorporating quadratic transformations into the models. A subgroup analysis was performed disaggregating the models by DMD exposure status, with separate models run for the group of patients exposed to DMD during follow-up and those non-exposed. A subgroup analysis test for interaction was used to assess for sub-group specific effects. All modelling analysis was adjusted for country of clinic contributing the patient data. All reported p-values are two-tailed and for each analysis p<0.05 was considered significant.
Prognostic nomogram

Using the independent prognostic correlates of second attack observed in the baseline adjusted modelling, we derived a predictive nomogram for conversion to CDMS using the method described by Katten et al. The nomogram was internally validated via derivation of the concordance index and nomogram calibration. The concordance index captures the probability that a MS patient drawn randomly from our dataset who was known to convert to CDMS before another randomly drawn patient, actually records a higher conversion probability on the prognostic nomogram. The index was derived by taking 1500 bootstrapped random samples of the original 3296 patients used to derive the multivariable “baseline only” Cox model described above. Another round of 1500 bootstrapped resamples was used to calibrate the nomogram. Patients were grouped according to their nomogram-predicted conversion probabilities and the means of these probability groups then compared against the empirically observed Kaplan-Meier conversion estimates on a calibration curve. These calibration curves represent the agreement between observed and predicted values across a range of predicted conversion probabilities. All analyses were conducted in Stata version 13 (StataCorp, College Station, TX) and R (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Patient characteristics

A total of 3296 patients from 50 clinics across 22 countries were eligible for this analysis (for patient characteristics, see Table 1), contributing a total of 5378.70 person-years of data. Three-hundred and eighty nine (42.7% of total initiates) commenced IM-IFNβ-1a, 308 (33.8%) SC-IFNβ-1a, 167 (18.4%) IFNβ-1b and 125 (13.7%) glatiramer acetate. Of the eligible patients, 1953 (59.3%) recorded a first post-CIS relapse event during follow-up. Incidence of first post-CIS relapse varied markedly by whether a patient was exposed to DMD therapy during observation or not, with the exposed subset
recording an incidence of 17.1 relapses per 100 person-years (95% CI 15.6, 18.8) compared with 53.8 per 100 person-years in the non-exposed group.

Table 1 – Demography, disease, treatment and examination characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Level</th>
<th>All (n=3296)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second events - n (%)</td>
<td>-</td>
<td>1953 (59.3)</td>
</tr>
<tr>
<td>Cumulative follow-up(^a) - person-years</td>
<td>-</td>
<td>5378.7</td>
</tr>
<tr>
<td>Time to second event (years)</td>
<td>Mean (SD)</td>
<td>1.09 (1.42)</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>0.50 (0.23, 1.27)</td>
</tr>
<tr>
<td>Gender - n (%)</td>
<td>Female</td>
<td>2324 (70.5)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>972 (29.5)</td>
</tr>
<tr>
<td>Age at MS onset - median (IQR)</td>
<td>-</td>
<td>31.61 (25.26, 39.33)</td>
</tr>
<tr>
<td>EDSS at CIS - median (IQR)</td>
<td>-</td>
<td>2 (1, 2.5)</td>
</tr>
<tr>
<td>Medication Possession Ratio (MPR)(^a) - mean (SD)</td>
<td>-</td>
<td>0.18 (0.33)</td>
</tr>
<tr>
<td>DMD exposed</td>
<td>-</td>
<td>910 (27.6)</td>
</tr>
<tr>
<td>Neuroanatomical location of first symptoms</td>
<td>Optic pathways</td>
<td>713 (21.6)</td>
</tr>
<tr>
<td></td>
<td>Supratentorial</td>
<td>675 (20.5)</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>710 (21.5)</td>
</tr>
<tr>
<td></td>
<td>Spinal Cord</td>
<td>847 (25.7)</td>
</tr>
<tr>
<td>Baseline Cerebral MRI</td>
<td>1+ T1 gadolinium enhancing lesion</td>
<td>571 (17.3)</td>
</tr>
<tr>
<td></td>
<td>9+ T2 hyperintensive lesions</td>
<td>909 (27.6)</td>
</tr>
<tr>
<td></td>
<td>1+ infratentorial lesion</td>
<td>1246 (37.8)</td>
</tr>
<tr>
<td></td>
<td>1+ juxtacortical lesion</td>
<td>1479 (44.9)</td>
</tr>
<tr>
<td></td>
<td>2+ periventricular lesions</td>
<td>1547 (46.9)</td>
</tr>
<tr>
<td>Baseline Spinal MRI(^b)</td>
<td>1+ T1 gadolinium enhancing lesion</td>
<td>49 (1.5)</td>
</tr>
<tr>
<td></td>
<td>1+ T2 hyperintensive lesion</td>
<td>535 (16.2)</td>
</tr>
<tr>
<td>Baseline CSF(^c)</td>
<td>Oligoclonal bands detected</td>
<td>1059 (32.1)</td>
</tr>
</tbody>
</table>

EDSS = Expanded Disability Status Scale, CIS = clinically isolated syndrome, DMD = Disease-modifying drug, MRI = magnetic resonance imaging, CSF = cerebrospinal fluid

a. The medication possession ratio is the proportion of follow-up time spent treated with a disease-modifying drug
b. Baseline spinal MRI was available for n=1539 patients
c. Baseline CSF examination was available for n=3104 patients. Of these n=1265 recorded an OCB status of absent or detected.

Predictors of conversion to CDMS

Older age at CIS was associated with a 10% reduction in the risk of conversion (5-year units, adjusted Hazard Ratio (aHR 0.90), 95% CI 0.88, 0.92; Table 2). Every point increase in baseline EDSS at CIS was associated with 1.16 times the rate of subsequent conversion (aHR 1.16, 95% CI 1.12, 1.20). Neuroanatomical location of first symptoms was also associated with rate of conversion, with a brainstem and supratentorial location associated with 1.17 (aHR 1.17, 95% CI 1.02, 1.36) and 1.29
(aHR 1.29, 95% CI 1.12, 1.48) times the rate of second attack, respectively, relative to an optic pathway location. Any exposure to DMD during follow-up was associated with a 42% rate reduction in time to first relapse compared with DMD-naïve patients (aHR 0.58, 95% CI 0.46, 0.73) (Figure 1) whilst a 1 unit increase in MPR was associated with a 65% reduction in the rate of relapse (aHR 0.35, 95% CI 0.25, 0.49).

Table 2 – Predictors of first post-CIS relapse

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Level</th>
<th>uHR (95% CI) p-value</th>
<th>aHR (95% CI) p-value&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>aHR (95% CI) p-value&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>1.10 (1.00, 1.22)</td>
<td>1.05 (0.95, 1.16) 0.338</td>
<td>1.12 (1.01, 1.23) 0.031</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Age at MS onset (5 years)</td>
<td>-</td>
<td>0.93 (0.91, 0.95)</td>
<td>0.90 (0.88, 0.92) &lt;0.001</td>
<td>0.92 (0.90, 0.94) &lt;0.001</td>
</tr>
<tr>
<td>EDSS (continuous)</td>
<td>-</td>
<td>1.13 (1.10, 1.17)</td>
<td>1.16 (1.12, 1.20) &lt;0.001</td>
<td>1.15 (1.11, 1.19) &lt;0.001</td>
</tr>
<tr>
<td>Exposed to DMD during follow up</td>
<td>-</td>
<td>0.37 (0.33, 0.42)</td>
<td>0.58 (0.46, 0.73) &lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Medication possession ratio (MPR)</td>
<td>-</td>
<td>0.24 (0.20, 0.28)</td>
<td>0.35 (0.25, 0.49) &lt;0.001</td>
<td></td>
</tr>
<tr>
<td>First symptoms location</td>
<td>Optic pathways</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Supratentorial</td>
<td>1.40 (1.22, 1.61)</td>
<td>1.29 (1.12, 1.48) 0.001</td>
<td>1.36 (1.18, 1.56) &lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>1.30 (1.14, 1.50)</td>
<td>1.17 (1.02, 1.36) 0.030</td>
<td>1.20 (1.04, 1.38) 0.015</td>
</tr>
<tr>
<td></td>
<td>Spinal Cord</td>
<td>1.20 (1.05, 1.37)</td>
<td>1.10 (0.96, 1.27) 0.179</td>
<td>1.15 (1.00, 1.31) 0.054</td>
</tr>
<tr>
<td>T1 Gadolinium lesions</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.18 (1.04, 1.33)</td>
<td>1.24 (1.09, 1.41) 0.001</td>
<td>1.09 (0.96, 1.24) 0.189</td>
</tr>
<tr>
<td>T2 hyperintensive lesions</td>
<td>0-2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>3-8</td>
<td>1.22 (1.04, 1.44)</td>
<td>0.98 (0.81, 1.20) 0.873</td>
<td>1.03 (0.85, 1.25) 0.760</td>
</tr>
<tr>
<td></td>
<td>9+</td>
<td>1.36 (1.14, 1.62)</td>
<td>0.96 (0.77, 1.20) 0.714</td>
<td>0.99 (0.79, 1.23) 0.909</td>
</tr>
<tr>
<td>Infratentorial lesions</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.36 (1.23, 1.51)</td>
<td>1.21 (1.08, 1.36) 0.001</td>
<td>1.15 (1.03, 1.29) 0.013</td>
</tr>
<tr>
<td>Juxta cortical lesions</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.21 (1.09, 1.34)</td>
<td>1.21 (1.06, 1.37) 0.004</td>
<td>1.03 (0.91, 1.17) 0.654</td>
</tr>
<tr>
<td>Periventricular lesions</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1-2</td>
<td>1.09 (0.91, 1.31)</td>
<td>1.15 (0.95, 1.40) 0.157</td>
<td>1.04 (0.86, 1.25) 0.817</td>
</tr>
</tbody>
</table>
The presence of CSF-restricted OCB was associated with 1.52 times the rate of relapse compared with OCB absence (aHR 1.52, 95% CI 1.22, 1.88). Baseline cerebral and spinal MRI lesion number and distribution were also associated with first post-CIS relapse. At least 1 T1 gadolinium enhancing lesion was associated with 1.24 times the rate of relapse (aHR 1.24, 95% CI 1.09, 1.41) whilst 3 or more periventricular lesions were associated with 1.68 times the rate of relapse compared with zero.
lesions. At least 1 infratentorial and at least 1 juxtacortical lesion on cerebral MRI were associated with 1.21 times (aHR 1.21, 95% CI 1.08, 1.36) and 1.21 times (aHR 1.21, 95% CI 1.06, 1.37) the rate of first post-CIS relapse, respectively, when compared with zero lesions. A greater number of T2 hyperintensive lesions, although associated on unadjusted modelling, were not predictive on adjusted modelling. Spinal T2 lesions were excluded from the adjusted modelling secondary to collinearity with the presence of OCB. As an additional sensitivity analysis we re-analysed the adjusted model substituting spinal T2 lesions for the presence of OCB predictor (Supplementary table 1). No significant interactions between any pairwise combinations of concurrent explanatory variables were identified in any of the adjusted models. There was no evidence of non-linear relationships between any of the explanatory variables tested and the dependent outcome variable and thus untransformed independent variables were used in the models. A sensitivity analysis limiting the primary model to those patients recording a baseline MRI within 3 months of baseline returned similar results with older age at CIS, higher EDSS, a supratentorial symptoms location, one or more infratentorial lesions and 3 or more periventricular lesions all again correlating with a higher rate of conversion (SupplementaryTable 2).

Prognostic nomogram

When the primary model was limited to baseline characteristics only (i.e. only factors recorded at CIS, excluding post-CIS DMD exposure metrics) age, female sex, EDSS, neuroanatomical location of first symptoms, T1 Gd+ infratentorial and periventricular lesions and the presence of OCB in baseline CSF again correlated with subsequent conversion (Table 2). Unlike the primary model, T1 Gd+ lesions were not associated with conversion whilst female sex now correlated with an increased conversion rate (aHR 1.12, 95% CI 1.01, 1.23). Using the significant baseline correlates of conversion identified in this adjusted Cox model, we derived a nomogram to predict 12 month conversion (Figure 2). The degree of contribution of each independent explanatory variable to the 12-month conversion
nomogram points were, in descending order: EDSS at CIS, age, first symptom neuroanatomical location, periventricular lesions, the presence of CSF oligoclonal bands, infratentorial lesions and female sex. The concordance index for the 12-month conversion model was 0.81. Additional nomograms for 6-month, 2-, 3-, 4- and 5-year conversion are presented as supplementary figures (supplementary figures 2-6). The concordance indices for the 6 month, 2-, 3-, 4-and 5-year conversion models were 0.76, 0.81, 0.82, 0.83 and 0.83 respectively. Supplementary figure 1 provides a worked example of how to interpret the nomogram using a hypothetical patient. The calibration curve (Figure 3) illustrates how the nomogram predictions compared with the actual observed cohort outcomes.

**Figure 2 – Nomogram for 12-month conversion.** To calculate total points for a patient, match response for a predictor to the top points scale, repeat for all prognostic factors and sum to derive total points. Match position on the total points scale to the conversion probability scale to identify the individualised probability of 12-month conversion.
Supplementary Figure 1 – Worked example of how to interpret nomogram. First match each predictor variable with a number of points on the top “Points” scale. For example female sex (black arrow) matches to 8 points on the “Points” scale, an age at CIS of 30 years (red arrow) matches to 62 points and an EDSS of 2 (grey arrows) to 20 points. Repeat for all 7 predictor variables and sum to obtain the total points. Thus for a female patient aged 30 at CIS with an EDSS of 2, a supratentorial first symptoms location, zero infratentorial and 1-2 periventricular lesions on cerebral MRI and oligoclonal bands detected on CSF examination would score a total of 138 points (blue arrow on the “Total Points” scale). Drawing a line down from the “Total Points” scale to the “Probability of 12-month conversion” scale demonstrates that 138 points corresponds to a 12-month conversion probability of 0.53 (or 53%) for this particular patient (green arrow).
Figure 3 – Calibration curve for 12-month conversion. Nomogram-predicted conversion probability is represented on the x-axis while the y-axis gives the actual 12-month conversion probability by the Kaplan-Meier method. Perfect agreement between actual and predicted conversion probability is represented by the blue-grey diagonal line.

**Subgroup analysis**

A subgroup analysis of patients who reported 1) DMD exposure during follow-up or 2) remained untreated during follow-up showed differences in the both the pattern of the predictors and the magnitude of their associations (Table 3). Many of the associations observed at the level of the entire sample decreased in magnitude and/or level of significance when limited to the DMD exposed subgroup only. By contrast these associations were generally larger and more significant in the non-DMD exposed subgroup. By extending the model to test formally for an interaction between DMD exposure and MRI metric we observed a significant difference in the association between MRI metrics and second attack rate by DMD exposure group (all p<0.001), with the exception of T2 hyperintensive lesions (p=0.438). There was no difference in baseline EDSS between the DMD exposed and non-exposed groups (median; IQR: 2; 1-1.5 in both groups; p=0.112).
Table 3 - Predictors of first post-CIS relapse – by DMD exposure status

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Exposed to DMD during follow up (n=910)&lt;sup&gt;a,b,d&lt;/sup&gt;</th>
<th>Not exposed to DMD during follow up (n=2386)&lt;sup&gt;a,c,d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Second event - n=438</td>
<td>Second event - n=1515</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>aHR (95% CI) p-value&lt;sup&gt;b&lt;/sup&gt;</td>
<td>aHR (95% CI) p-value&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Age at MS onset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0.92 (0.87, 0.97) 0.002</td>
<td>0.91 (0.89, 0.93) 0.001</td>
</tr>
<tr>
<td>EDSS (continuous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>1.22 (1.10, 1.34) &lt;0.001</td>
<td>1.16 (1.12, 1.21) &lt;0.001</td>
</tr>
<tr>
<td>First symptoms location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optic pathways</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Supratentorial</td>
<td>1.24 (0.84, 1.83) 0.272</td>
<td>1.28 (1.08, 1.52) 0.005</td>
</tr>
<tr>
<td>Brainstem</td>
<td>0.89 (0.59, 1.34) 0.575</td>
<td>1.30 (1.10, 1.55) 0.003</td>
</tr>
<tr>
<td>Spinal Cord</td>
<td>0.74 (0.49, 1.11) 0.145</td>
<td>1.18 (1.00, 1.40) 0.054</td>
</tr>
<tr>
<td>T1 Gadolinium lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1+</td>
<td>1.44 (1.08, 1.91) 0.012</td>
<td>1.29 (1.11, 1.50) 0.001</td>
</tr>
<tr>
<td>T2 hyperintensive lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3-8</td>
<td>1.24 (0.65, 2.35) 0.514</td>
<td>1.19 (0.61, 2.10) 0.489</td>
</tr>
<tr>
<td>9+</td>
<td>1.18 (0.60, 2.32) 0.628</td>
<td>1.15 (0.58, 2.18) 0.397</td>
</tr>
<tr>
<td>Infratentorial lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1+</td>
<td>1.23 (0.95, 1.58) 0.116</td>
<td>1.29 (1.14, 1.47) &lt;0.001</td>
</tr>
<tr>
<td>Juxtacortical lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1+</td>
<td>1.15 (0.86, 1.54) 0.354</td>
<td>1.17 (1.01, 1.34) 0.032</td>
</tr>
<tr>
<td>Periventricular lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1-2</td>
<td>0.93 (0.49, 1.75) 0.818</td>
<td>0.96 (0.63, 1.51) 0.703</td>
</tr>
<tr>
<td>3+</td>
<td>1.00 (0.53, 1.90) 0.991</td>
<td>1.70 (1.39, 2.08) &lt;0.001</td>
</tr>
<tr>
<td>Oligoclonal bands in CSF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Present</td>
<td>1.27 (0.68, 2.37) 0.454</td>
<td>1.40 (1.11, 1.77) 0.004</td>
</tr>
</tbody>
</table>

a) Additionally adjusted for country
b) Hazard proportionality test: p=0.3688
c) Hazard proportionality test: p=0.4555
d) test for interaction: p=0.002

EDSS = Expanded Disability Status Scale, CIS = clinically isolated syndrome, DMD = Disease-modifying drug, MRI = magnetic resonance imaging, CSF = cerebrospinal fluid, aHR = adjusted Hazard ratio

DISCUSSION

This multinational, prospective study represents the largest post-CIS cohort reported to date, examining predictors of time to first relapse in 3296 CIS patients. We confirmed that, in a multivariable model, younger age at CIS onset, and increased baseline EDSS scores were clinical
predictors of shorter time to relapse, whereas sex had no independent effect. Each of the Barkhof-Tintore criteria (3+ periventricular lesions, 1+ infratentorial lesions, 1+ juxtacortical lesions) retained independent predictive power. Total cerebral T2 lesion number was not predictive of time to relapse attack, whereas the presence of 1+ cerebral Gd-enhancing lesion was predictive. Importantly, presence of oligoclonal bands was an independent predictor of shorter time to relapse, but this was highly collinear with the presence of 1+ spinal cord lesions. DMD exposure was strongly protective against relapse, consistent with results of several phase three trials. These results corroborate and extend prior, albeit smaller, studies observing similar sets of predictors of conversion probability.\textsuperscript{10,19,29} The recent Tintore study, based on follow-up in a single centre on 1015 patients, reports younger age at onset, presence of oligoclonal band and more than 10 brain MRI lesions as prognostic factors for developing MS.\textsuperscript{23} We report similar results on demographic characteristics and biological results and extend these to observe a sex effect in the “baseline factors only” model. Our larger multicentre study provides greater power to assess the predictive value of individual brain MRI criteria, with each of the Barkhof-Tintore criteria confirmed as independent factors shortening time to second attack.

Identification of patient, disease and examination factors associated with higher probability of second attack in clinical practice may enable clinicians to flag patients that could benefit from more intensive follow-up and consideration of early DMD treatment intervention, facilitating more favourable patient outcomes. The observation that, across both full and “baseline factors only” models the lesion location covariates out-perform lesion count metrics confirms the finding that cerebral lesion location is preferable to lesion number in calculating risk rate of second attack, as demonstrated by Swanton et al, (2007), and incorporated into the 2010 revisions of the McDonald criteria for diagnosis of MS.\textsuperscript{26,29}
Across both the primary and subgroup analyses, any exposure to DMD appears to be a more important and consistent predictor of first post-CIS outcome than the precise amount of exposure. These results support early initiation of DMD, particularly for those subsets that demonstrate overall higher probability of first post-CIS relapse (younger age at CIS, higher number of cerebral MRI lesions in specific locations, and lesion presence on spinal MRI or OCB presence in CSF). Furthermore the precise identity of the DMD product initiated did not change the effect greatly. This result is consistent with observations from the recent phase 3 REFLEX trial comparing two dosing frequencies of subcutaneous IFNβ-1a in first clinical demyelinating event patients which observed that either treatment delayed subsequent clinical relapse post-CIS, regardless of the dosing regimen used.30

The high correlation between the T2 hyperintensive lesions on spinal MRI with presence of OCB on baseline CSF examination could suggest a role for possible rationalisation of the battery of investigations around the time of onset. While greater lesion loads on cerebral MRI in patients with CIS has been correlated with a higher probability of finding OCB in concurrently sampled CSF,18,20 our results are novel in that the concurrence of spinal lesions and OCB presence is so strong that a multivariable statistical model cannot be performed with both predictors present, due to collinearity. It is reasonable to hypothesize that there is a biological explanation- either OCB’s (i.e. oligoclonal antibodies) could be produced preferentially in or near spinal lesions, or OCB’s could be causally involved in the generation of spinal lesions.

The nomogram permits calculation of the cumulative effect of multiple prognostic factors of individualised conversion probability. By weighting the influence of each factor, the nomogram provides an appreciation of the relative magnitude of influence of each prognostic factor on conversion probability. Younger age and higher EDSS at CIS dominated the nomogram in terms relative contribution to total points and subsequent conversion probability, consistent with prior literature detailing risk factors for conversion. The advantage of a nomogram over an adjusted
regression model is that, whilst the latter returns estimates of the average effects across a population, the nomogram permits individualised predictions to be made, which may be useful for both patient management and potentially inform future revisions of diagnostic criteria, which are typically based on predictors of time to clinical relapse. Whilst our own internal validation suggested good performance, external validation through the application of the nomogram to a separate MS dataset or population is required to confirm the generalisability of the nomogram.

This large post-CIS prospective study describes the independent predictive effect of demographic, clinical, radiological and CSF predictors of time to clinically definite MS. Our results confirm and extend those of many prior studies and validate the use of predictors in clinical practice. The characterisation of independent predictors allows more accurate prognosis for CIS patients and could inform future revisions of the diagnostic criteria of relapsing-remitting MS.

**Funding**

The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBasis study was specifically supported by Merck Serono, between 2004 and 2009. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. Claire Meyniel was supported by the Journées de Neurologie de Langue Française.

**Author contributions:**

**T Spelman** conceptualised and designed the study, conducted and interpreted the analysis and drafted, revised and approved the manuscript.

**C Meyniel** conceptualised and designed the study, interpreted the analysis and drafted, revised and approved the manuscript.
H Butzkueven conceptualised the study, interpreted the analysis and revised and approved the manuscript.

JI Rojas, T Kalincik and V Jokubaitis, interpreted the analysis, revised and approved the manuscript.

A Lugaresi, G Izquierdo, F Grand’Maison, C Boz, R Alroughani, E Havrdova, D Horakova, G Iuliano, P Duquette, M Terzi, P Grammond, JA Cabrera-Gomez, R Hupperts, J Lechner-Scott, C Oreja-Guevara, R Fernández Bolaños, E Pucci, F Verheul, M Fiol, V Van Pesch, E Cristiano, T Petersen, T Petkovska-Boskova, F Moore, I Kister, and M Trojano contributed substantially to data acquisition, interpretation of the analysis and have revised and approved the manuscript.

Declaration of interests:

Tim Spelman received compensation for serving on scientific advisory boards, honoraria for consultancy and funding for travel from Biogen Inc; speaker honoraria from Novartis.

Claire Meyniel received travel grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva.

Juan Ignacio Rojas did not declare any competing interests.

Alessandra Lugaresi was a Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, travel and research grants from the Associazione Italiana Sclerosi Multipla.

Guillermo Izquierdo received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

Francois Grand-Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.
Timothy Denis Spelman (58172)

**Cavit Boz** received conference travel support from Biogen Idec, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.

**Raed Alroughani** received honoraria from Biologix, Biogen, Bayer, Genzyme, Genpharm, Merck Serono, GSK and Novartis, and served on advisory board for Bayer, Biologix, Biogen, Genzyme, Genpharm, Novartis and Merck Serono.

**Eva Havrdova** received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

**Dana Horakova** received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

**Gerardo Iuliano** received honoraria from Biogen-Idec, Novartis, Sanofi, Serono and Teva.

**Pierre Duquette** has received honoraria for organising CME events and has obtained funding to attend meetings from Biogen Idec, EMD Serono, TEVA Neuroscience, Novartis, and Genzyme, has received funding for investigator-initiated trials with Biogen Idec, EMD Serono and Novartis, and has received peer-review funding from CIHR and from the MS Society of Canada.

**Murat Terzi** received travel grants from Merck Serono, Novartis, Bayer-Schering and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.

**Pierre Grammond** is a Novartis, Teva-neuroscience, Biogen Idec advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis.

**Jose Antonio Cabrera-Gomez** did not declare any competing interests

**Raymond Hupperts** received honoraria as consultant on scientific advisory boards from Merck-Serono, Biogen-Idec, Genzyme-Sanofi and Teva, research funding from Merck-Serono and Biogen-Idec, and speaker honoraria from Sanofi-Genzyme.
Timothy Denis Spelman (58172)

**Jeannette Lechner-Scott** has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen, CSL, Genzyme Sanofi, Merck Serono, Novartis and TEVA.

**Celia Oreja-Guevara** received honoraria as scientific advisory board consultant from Biogen-Idec, Bayer-Schering, Merck-Serono, Teva and Novartis; has participated in research projects by Biogen-Idec, GSK, Teva and Novartis

**Ricardo Fernández Bolaños** did not declare any competing interests

**Eugenio Pucci** served on scientific advisory boards for Genzyme, Novartis and Biogen-Idec; he has received honoraria and travel grants from Sanofi Aventis, Novartis, Biogen Idec, Merck Serono, Genzyme and Teva; he has received travel grants from Associazione Marchigiana Sclerosi Multipla e altre malattie neurologiche.

**Freek Verheul** did not declare any competing interests.

**Marcela Fiol** did not declare any competing interests.

**Vincent Van Pesch** has received travel grants and honoraria for consultancy or lectures from Bayer-Schering, Biogen Idec, Merck Serono, Novartis Pharma, Sanofi-Aventis and Teva

**Edgardo Cristiano** did not declare any competing interests.

**Thor Petersen** did not declare any competing interests.

**Tatjana Petkovska-Boskova** has accepted travel compensation from Biogen.

**Fraser Moore** did not declare any competing interests.

**Ilya Kister** did not declare any competing interests.

**Tomas Kalincik** received compensation for conference travel and speaker honoraria from Novartis, Biogen Idec, Genzyme, Sanofi Aventis, Teva, BioCSL and Merck Serono and served on advisory boards for Novartis, Merck Serono and Biogen.

**Vilija Jokubaitis** received compensation for conference travel from Novartis
Timothy Denis Spelman (58172)

**Maria Trojano** received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

**Helmut Butzkueven** received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital.

**MSBasis study group co-investigators and contributors:** From the MS-Centrum Nijmegen, Nijmegen, The Netherlands, Dr Cees Zwanikken; from Flinders Medical Centre, Adelaide, Australia, Dr Mark Slee; from Geelong Hospital, Geelong, Australia, Dr Cameron Shaw; from the Brain and Medical Research Institute, University of Sydney, Sydney, Australia, Dr Michael Barnett; from St Vincent’s Hospital, Fitzroy, Australia, Dr Neil Shuey; from C. Mondino National Neurological Institute, Pavia, Italy Dr Roberto Bergamaschi; from Craigavon Area Hospital, Portadown, United Kingdom, Dr Stella Hughes; from INEBA, Buenos Aires, Argentina, Dr Maria Laura Saladino; from Mater Dei Hospital, Malta; Dr Norbert Vella; from Department NEUROFARBA, Section of Neurosciences, University of Florence, Florence, Italy, Dr Maria Pia Amato; from Hospital Tenon, Paris, France, Dr Etienne Roullet; from UT Southwestern, Dallas, United States, Dr Benjamin Greenberg; from Central Clinical Emergency Military Hospital, Bucharest, Romania, Dr Carmen-Adella Sirbu; from Hospital Angeles de las Lomas. Instituto Mexicano de Neurociencias, Huixquilucan de Degollado, Mexico, Dr Eli Skromne; from University Clinic of Neurology, Skopje, Macedonia, Dr Vladimir Bojkovski; from Universita Napoli, Napoli, Italy, Dr Pietro Carrieri; from The Walton Centre for Neurology and Neurosurgery, Liverpool, United Kingdom, Dr Carolyn Young; from Hospital Donostia, Gipuzkoa, Spain, Dr Javier Olascoaga; from Consultorio Privado, Buenos Aires, Argentina, Dr Aldo Savino; and from Instituto de Neurociencias Cordoba, Cordoba, Argentina, Dr Elizabeth
Timothy Denis Spelman (58172)

Alejandra Bacile. No compensation has been received for the persons who have made substantial contributions to the work but do not qualify as authors.

REFERENCES


Supplementary Table 1 – Predictors of first post-CIS relapse – alternate model substituting T2 lesions on spinal MRI for oligoclonal bands on CSF examination

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Level</th>
<th>aHR (95% CI)</th>
<th>p-valuea,b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>1.04 (0.94, 1.15)</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Age at MS onset (5 years)</td>
<td>-</td>
<td>0.90 (0.88, 0.92)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>EDSS (continuous)</td>
<td>-</td>
<td>1.16 (1.12, 1.20)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Exposed to DMT during follow up</td>
<td>-</td>
<td>0.59 (0.47, 0.74)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medication possession ratio (MPR)</td>
<td>-</td>
<td>0.36 (0.26, 0.50)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>First symptoms location</td>
<td>Optic pathways</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supratentorial</td>
<td>1.29 (1.12, 1.48)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>1.17 (1.02, 1.36)</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>Spinal Cord</td>
<td>1.10 (0.96, 1.27)</td>
<td>0.179</td>
</tr>
<tr>
<td>McDonald T1 Gadolinium lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.31 (1.15, 1.50)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>McDonald T2 hyperintensive lesions</td>
<td>0-2</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-8</td>
<td>0.99 (0.81, 1.21)</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td>9+</td>
<td>0.97 (0.78, 1.21)</td>
<td>0.800</td>
</tr>
<tr>
<td>McDonald Infratentorial lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.30 (1.16, 1.46)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>McDonald Juxtacortical lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>1.17 (1.03, 1.33)</td>
<td>0.015</td>
</tr>
<tr>
<td>McDonald Periventricular lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-2</td>
<td>1.22 (0.99, 1.50)</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>3+</td>
<td>1.78 (1.46, 2.17)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Spinal MRI - number of T1 gadolinium lesions</td>
<td>-</td>
<td>1.23 (0.91, 1.66)</td>
<td>0.172</td>
</tr>
<tr>
<td>Spinal MRI - number of T2 lesions</td>
<td>-</td>
<td>1.35 (1.01, 1.77)</td>
<td>0.023</td>
</tr>
<tr>
<td>Oligoclonal bands in CSF</td>
<td>Absent</td>
<td>Co-linear with Spinal MRI lesions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>Co-linear with Spinal MRI lesions</td>
<td></td>
</tr>
</tbody>
</table>

a) Additionally adjusted for country
b) Hazard proportionality test: p=0.7819
Supplementary Table 2 – Sensitivity analysis: Predictors of first post-CIS relapse – limiting primary model to subset of patients with baseline MRI within 3 months of CIS

<table>
<thead>
<tr>
<th>Predictor at CIS</th>
<th>Level</th>
<th>aHR (95% CI)</th>
<th>p-value*α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.12</td>
<td>(0.99, 1.27)</td>
<td>0.067</td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at MS onset (5 years)</td>
<td>-</td>
<td>0.91</td>
<td>(0.88, 0.94)</td>
</tr>
<tr>
<td>EDSS</td>
<td>-</td>
<td>1.11</td>
<td>(1.06, 1.14)</td>
</tr>
<tr>
<td>First symptoms location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optic pathways</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supratentorial</td>
<td>1.27</td>
<td>(1.07, 1.50)</td>
<td>0.007</td>
</tr>
<tr>
<td>Brainstem</td>
<td>1.13</td>
<td>(0.95, 1.35)</td>
<td>0.162</td>
</tr>
<tr>
<td>Spinal Cord</td>
<td>1.18</td>
<td>(0.99, 1.39)</td>
<td>0.062</td>
</tr>
<tr>
<td>McDonald T1 Gadolinium lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1+</td>
<td>1.11</td>
<td>(0.95, 1.30)</td>
<td>0.181</td>
</tr>
<tr>
<td>McDonald T2 hyperintensive lesions</td>
<td>0-2</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3-8</td>
<td>0.97</td>
<td>(0.78, 1.22)</td>
<td>0.819</td>
</tr>
<tr>
<td>9+</td>
<td>0.96</td>
<td>(0.74, 1.25)</td>
<td>0.759</td>
</tr>
<tr>
<td>McDonald Infratentorial lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1+</td>
<td>1.19</td>
<td>(1.03, 1.37)</td>
<td>0.019</td>
</tr>
<tr>
<td>McDonald Juxtatentorial lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1+</td>
<td>1.00</td>
<td>(0.85, 1.17)</td>
<td>0.991</td>
</tr>
<tr>
<td>McDonald Periventricular lesions</td>
<td>0</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>0.97</td>
<td>(0.76, 1.22)</td>
<td>0.771</td>
</tr>
<tr>
<td>3+</td>
<td>1.37</td>
<td>(1.08, 1.74)</td>
<td>0.009</td>
</tr>
<tr>
<td>Spinal MRI - number of T1 gadolinium lesions</td>
<td>-</td>
<td>Co-linear with Oligoclonal bands</td>
<td></td>
</tr>
<tr>
<td>Spinal MRI - number of T2 gadolinium lesions</td>
<td>-</td>
<td>Co-linear with Oligoclonal bands</td>
<td></td>
</tr>
<tr>
<td>Oligoclonal bands present in CSF</td>
<td>-</td>
<td>1.20</td>
<td>(0.95, 1.52)</td>
</tr>
</tbody>
</table>

a) Adjusted for covary
b) Hazard proportionality test: p=0.4016
Supplementary Figure 2 – Nomogram for 6-month conversion

<table>
<thead>
<tr>
<th>Points</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>85</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>80</td>
<td>70</td>
<td>60</td>
<td>50</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDSS</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First symptom location</td>
<td>Optic Pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 Od+ Infratentorial Lesions</td>
<td>1+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 Od+ Periventricular Lesions</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oligoclonal bands in CSF</td>
<td>Absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Points</td>
<td></td>
<td>0</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
<td>180</td>
<td>200</td>
<td>220</td>
<td>240</td>
</tr>
<tr>
<td>Probability of 6-month conversion</td>
<td>0.15</td>
<td>0.20</td>
<td>0.25</td>
<td>0.30</td>
<td>0.35</td>
<td>0.40</td>
<td>0.45</td>
<td>0.50</td>
<td>0.55</td>
<td>0.60</td>
<td>0.65</td>
<td>0.70</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

Supplementary Figure 3 – Nomogram for 2-year conversion

<table>
<thead>
<tr>
<th>Points</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>85</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>80</td>
<td>70</td>
<td>60</td>
<td>50</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDSS</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First symptom location</td>
<td>Optic Pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 Od+ Infratentorial Lesions</td>
<td>1+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 Od+ Periventricular Lesions</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oligoclonal bands in CSF</td>
<td>Absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Points</td>
<td></td>
<td>0</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
<td>180</td>
<td>200</td>
<td>220</td>
<td>240</td>
</tr>
<tr>
<td>Probability of 2-year conversion</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure 4 – Nomogram for 3-year conversion

Supplementary Figure 5 – Nomogram for 4-year conversion
Supplementary Figure 6 – Nomogram for 5-year conversion
F. ANALYSIS: Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict early discontinuation

F.1 Gender, country, baseline EDSS and on-treatment relapse rate predict first DMT treatment discontinuation (paper-format report)

TITLE: Gender, country, baseline EDSS and on-treatment relapse rate predict first DMT treatment discontinuation

AUTHORS: Tim Spelman¹, Maria Trojano², Eva Havrdova³, Alessandra Lugaresi⁴, Guillermo Izquierdo⁵, Francois Grand’Maison⁶, Raed Alroughani⁷, Pierre Grammond⁸, Pierre Duquette⁹, Raymond Hupperts¹⁰, Cavit Boz¹¹, Murat Terzi¹², Celia Oreja-Guevara¹³, Jose Antonio Cabrera-Gomez¹⁴, Gerardo Iuliano¹⁵, Ricardo Fernández Bolaños¹⁶, Jeannette Lechner-Scott¹⁷, Helmut Butzkueven¹, on behalf of the MSBASIS (an MSBase substudy) Investigators

¹Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia

²Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy

³Department of Neurology and Center of Clinical Neuroscience, 1st Faculty of Medicine, General University Hospital and Charles University in Prague, Czech Republic

⁴MS Center, Department of Neuroscience and Imaging, University ‘G. d’Annunzio’, Chieti, Italy

⁵Hospital Universitario Virgen Macarena, Sevilla, Spain

⁶Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada

⁷Amiri Hospital, Kuwait City, Kuwait

⁸Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada

⁹Hôpital Notre Dame, Montreal, Canada
Timothy Denis Spelman (58172)

10 Maaslandziekenhuis, Sittard, The Netherlands
11 Karadeniz Technical University, Trabzon, Turkey
12 Mayis University, Samsun, Turkey
13 University Hospital San Carlos, Madrid, Spain
14 Centro Internacional de Restauracion Neurologica, Havana, Cuba
15 Ospedali Riuniti di Salerno, Salerno, Italy
16 Hospital Universitario Virgen de Valme, Seville, Spain
17 John Hunter Hospital, Newcastle, Australia

**Corresponding Author:** Tim Spelman

Postal Address:

Melbourne Brain Centre, Royal Melbourne Hospital,

Grattan St, Parkville, Victoria, Australia 3050

Ph: +61 3 9342 8070

Fax: +61 3 9342 8070

Email: tim@burnet.edu.au
Abstract

Background: Treatment discontinuation is common in MS. Available disease-modifying drugs (DMD) mitigate disease but are only partially effective and on-treatment breakthrough disease is a common occurrence and treatment discontinuation is associated with poorer health outcomes. Independent predictors of first DMD discontinuation have not previously been examined using a prospectively observed and seen-from-disease-onset cohort of this size and global cover.

Methods: DMD commencements were sourced from the MSBase Incident Study (MSBASIS) subset of the MSBase global registry. Predictors of time to first treatment discontinuation were analysed using a Cox proportional hazards regression and predictors of time to any treatment discontinuation were studied using a conditional risk-set model.

Results: 2003 first DMT initiations were assessed in this wholly seen-from-onset cohort. Females were associated with 1.46 times the rate of discontinuation compared with males (HR 1.46, 95% CI 1.21, 1.77, p<0.001). Every 1 unit increase in baseline EDSS was associated with 1.1 times the rate of subsequent treatment discontinuation (HR 1.10, 95% CI 1.03, 1.18, p=0.003) whilst every 1 relapse per person-year increase in ARR was associated with 1.35 times the rate of discontinuation (HR 1.35, 95% CI 1.30, 1.41, p<0.001). Patients managed within Italian, Canadian, Spanish, Czech and Turkish clinics were associated with 55%, 60%, 59%, 72% and 43% reductions in the rate of discontinuation compared with patients managed in Australian clinics whilst first IM-IFNβ-1a was associated with an increased risk of discontinuation compared with either SC-IFNβ-1a, IFNβ-1b or glatiramer acetate. On-treatment EDSS change, age at either disease onset or first DMT initiation and baseline cerebral MRI parameters were not association with discontinuation.

Conclusions: This large multinational observational study demonstrates that gender, clinic country, DMT identity, baseline EDSS and on-treatment ARR are all independently associated with the
probability and timing of first DMT discontinuation. Identification of patient and disease characteristics that are associated with higher probability of treatment discontinuation in clinical practice may enable the clinician to flag subsets of patients that may benefit from closer, more intensive follow-up whilst on treatment, facilitating more timely interventions and ultimately, more favourable patient outcomes.

**Keywords:** discontinuation, persistence, disease-modifying therapy, multiple sclerosis, relapse, EDSS, gender
Introduction

Treatment discontinuation is common in MS. Available disease-modifying drugs (DMD) mitigate disease but are only partially effective and on-treatment breakthrough disease is a common occurrence. Treatment discontinuation is associated with poorer health outcomes, including significantly higher EDSS and significantly lower proportion of patients remaining relapse and progression-free compared to patients who remained on DMD therapy.

The paucity of appropriately-powered, prospectively observed, real-world data around the characteristics, either demographic or disease-activity based, of patients who cease treatment makes identification and, importantly, anticipation of patient subsets likely to demonstrate sub-optimal response to currently available platform therapies difficult. This is particularly important early on in a patient’s treatment course where discontinuation rates have previously been observed to be as high as 20% within the first 6 months of DMD therapy and up to 70% within the first year of treatment.

Previous studies have identified a diverse range of patient, disease and treatment characteristics that correlate with DMD treatment discontinuation including gender, country, on-treatment relapse activity and EDSS change, disease duration and education level. Further, differences in persistence have been previously observed between different immunomodulatory therapies. However the performance of these characteristics as reliable and replicable predictors of treatment persistence have not before been examined in a real-world, prospectively assessed and seen-from-onset cohort of this size and global spread previously.

The objective of the current study was to examine pre- and on-treatment disease activity markers and demography as predictors of first DMD persistence using a large, multi-national, prospective,
observational registry cohort. To our knowledge, this is the largest such study to prospectively assess patients in clinical practice from disease onset.

**MATERIALS & METHODS**

**MSBase Registry**

The MSBase Registry is an international online database accumulator that was established in 2004 and collects disease related information from consenting patients attending Multiple Sclerosis (MS) clinics. The registry is a collaborative research group that prospectively collects outcomes data from MS treatment centres using an internet-based, physician owned and operated system www.msbase.org. Each center enters patient data in the offline iMed© local electronic database during routine clinic visits and intermittently uploads anonymized datasets to the MSBase server. Physicians record clinical information such as date of MS onset, Kurtzke Expanded Disability Status Score (EDSS), relapse characteristics, MRI and other investigations and diagnostic criteria used. Records are classified as complete and eligible for analyses if they meet a minimum required set of data. Quality of the EDSS assessment was assured by the requirement of online Neurostatus certification at each of the participating centres. The MSBase registry was approved by the Melbourne Health Human Research Ethics Committee, and by the local ethics committees in all participating centres (or exemptions granted, according to applicable local laws and regulations). Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing center.

**MSBASIS**

The MSBase Incident Study (MSBASIS) is a sub-study of the MSBase registry. Commencing on the 18th November 2004, MSBASIS is an ongoing global, longitudinal, investigator-initiated and maintained observational cohort study designed to prospectively assess all registry MS patients with
a clinically isolated syndrome (first demyelinating event) starting less than 12 months from the enrolment date. As of the 2nd April 2014, the date of data extract and compilation, the study had enrolled 4313 patients from 50 clinics across 22 countries contributing 37,569 clinic visit observation points.

Inclusions
MSBASIS requires a minimum entry dataset at each clinic visit observation point consisting of the visit date, EDSS, Kurtze functional systems score (KFS), onset date and duration of prospectively observed relapses in addition to any glucocorticoid therapy and, where applicable, initiation and discontinuation dates for disease-modifying drugs (DMD) used specifically for MS therapy. Patients were eligible for this analysis if CIS was diagnosed or confirmed by a participating neurologist within 12 months of the CIS onset date. Following this initial assessment, minimum annual follow-up was required although details of all visits within any given year of follow-up were recorded and used in this analysis. Full EDSS and KFS assessment, determined according to the Neurostatus system,14 were required at this initial assessment to qualify for this analysis. In addition a cerebral magnetic resonance imaging (MRI) scan using Barkhof criteria for lesion dissemination in space within 12 months of CIS onset was further required to be eligible.15-17 Patients with primary progressive MS (PPMS) were excluded. Of the original 4313 patients enrolled in MSBASIS, a total of 3299 patients satisfied these inclusion criteria and were thus eligible for this study.

Outcome measure & definitions
The primary outcome of this analysis was time to first DMT treatment discontinuation. First DMTs included in the analysis were IM-IFNβ-1a, SC-IFNβ-1a, IFNβ-1b, glatiramer acetate, natalizumab and fingolimod. Baseline was defined as the date of first DMT initiation following confirmed CIS onset. Baseline EDSS was defined as the EDSS recorded nearest to the baseline date within 6 months either side of the baseline date. EDSS change was calculated as the EDSS at either the point of censoring or
end of follow-up as appropriate, minus the baseline EDSS. Annualised relapse rate (ARR) was defined as the number of prospectively observed on-treatment relapses divided by the years of follow-up for a given patient.

**Statistical analyses**

We analysed, as a priori predictors, gender, age (at both disease onset and first DMT initiation), country, baseline EDSS, on-treatment EDSS change, on-treatment ARR and baseline MRI for association with time to first DMT discontinuation. Categorical variables were summarised using frequency and percentage. Continuous variables were assessed for significant departure from normality using a Shapiro-Wilk test and summarised using mean and standard deviation (SD) or median and inter-quartile range (IQR) as appropriate. DMT discontinuation rates were expressed as counts of discontinuation events per 100 person-years of follow-up. Kaplan-Meier survival curves were used to describe cumulative survival over the observation period. Cox proportional hazards regression was used to investigate correlation between our a priori identified predictors and time to first DMT treatment discontinuation. Patients were either censored at the date of first treatment discontinuation or, where no discontinuation event was observed, the date of the last recorded study visit per patient. All a priori predictors were included in the adjusted model except where two or more related explanatory variables demonstrated significant co-linearity or overlap such as ARR (all relapses) and ARR (steroid-treated relapses only) or baseline EDSS expressed as both a continuous and categorical variable. In these instances, the investigators decide between these highly correlate characteristics on the basis of both the predictor’s performance on unadjusted modelling in addition to assessment of clinical relevance and interpretability. Interactions between predictor variables in the adjusted model were tested for. Hazard proportionality was assessed through analysis of scaled Schoenfeld residuals. Predictors of time to any DMT discontinuation were analysed using a conditional risk-set model, appropriate for modelling time to multiple or recurrent discontinuation events per patient. This is similar to the simpler Andersen & Gill extension
of the Cox model for multiple event data but further includes information regarding the explicit ordering of DMT discontinuation events over a patient's follow-up and thus permits assessment of whether earlier discontinuations, be it the actual time taken to discontinue or the specific DMT product identity, influence or inform subsequent discontinuations.\textsuperscript{22-24} Hazard proportionality was again assessed through analysis of scaled Schoenfeld residuals. All reported p-values are two-tailed and for each analysis $p<0.05$ was considered significant. All analyses were conducted in Stata version 13 (StataCorp, College Station, TX) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Of the 3299 patients who satisfied the inclusion criteria as at the date of data extract and compilation (2\textsuperscript{nd} April 2014), 2003 commenced a first post-CIS onset DMT (Table 1) contributing a cumulative 4,457 person-years of follow-up. 1412 (70.5\%) of these patients were female. 453 (22.6\%) were followed up in Italian clinics, 280 (14.0\%) Canadian, 242 (12.1\%) Australian and 227 (11.3\%) in Spanish clinics. Median (IQR) age of CIS onset was 30.4 years (24.4, 37.7). Baseline EDSS was similar across all first DMD products with all recording a median baseline EDSS of 2 with the exception of IM-IFNβ-1a and fingolimod (1.5). Across all first DMT, 391 (19.5\%) recorded at least 1 T1 gadolinium enhancing lesion on baseline cerebral MRI, 641 (32.0\%) recorded nine or more T2 hyperintensive lesions, 1095 (54.7\%) at least 2 periventricular lesions and 861 (43.0\%) and 1006 (50.2\%) at least one infratentorial and juxtacortical lesion respectively.
Table 1 – Demography, disease and treatment characteristics

<table>
<thead>
<tr>
<th>Characteristic Level</th>
<th>All first DMD</th>
<th>SC-IFNβ-1a</th>
<th>IM-IFNβ-1a</th>
<th>IFNβ-1b</th>
<th>Glatiramer Acetate</th>
<th>Natalizumab</th>
<th>Fingolimod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients - n</td>
<td>2003</td>
<td>612</td>
<td>615</td>
<td>380</td>
<td>305</td>
<td>54</td>
<td>37</td>
</tr>
<tr>
<td>Discontinuations - n (%)</td>
<td>1076 (53.7)</td>
<td>358 (58.5)</td>
<td>342 (55.6)</td>
<td>207 (54.5)</td>
<td>147 (48.2)</td>
<td>16 (29.6)</td>
<td>6 (16.2)</td>
</tr>
<tr>
<td>Cumulative follow-up - person-years</td>
<td>4546.8</td>
<td>1269.6</td>
<td>1601.8</td>
<td>894.7</td>
<td>648.8</td>
<td>82.6</td>
<td>49.3</td>
</tr>
<tr>
<td>Time on first treatment (years) - mean (SD)</td>
<td>2.3 (2.0)</td>
<td>2.1 (2.0)</td>
<td>2.6 (2.1)</td>
<td>2.4 (1.9)</td>
<td>2.1 (2.0)</td>
<td>1.5 (1.2)</td>
<td>1.3 (1.3)</td>
</tr>
<tr>
<td>Time to discontinuation (years) - mean (SD)</td>
<td>1.8 (1.6)</td>
<td>1.7 (1.8)</td>
<td>1.9 (1.6)</td>
<td>1.8 (1.6)</td>
<td>1.7 (1.6)</td>
<td>1.8 (1.3)</td>
<td>0.6 (0.2)</td>
</tr>
<tr>
<td>Time from CIS onset to first DMT (months) - median (IQR)</td>
<td>7.2 (3.7, 13.3)</td>
<td>7.8 (4.3, 13.2)</td>
<td>5.7 (2.8, 11.0)</td>
<td>7.5 (3.8, 14.6)</td>
<td>8.4 (4.4, 15.9)</td>
<td>6.5 (3.9, 15.2)</td>
<td>17.1 (4.9, 59.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender - n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country - n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
</tr>
<tr>
<td>Italy</td>
</tr>
<tr>
<td>Canada</td>
</tr>
<tr>
<td>Spain</td>
</tr>
<tr>
<td>Netherlands</td>
</tr>
<tr>
<td>Czech Republic</td>
</tr>
<tr>
<td>Turkey</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

| Age at MS onset - median (IQR) | 30.4 (24.4, 37.7) | 29.6 (23.6, 37.6) | 30.1 (24.2, 37.1) | 30.7 (25.3, 37.9) | 31.7 (26.9, 39.3) | 25.9 (19.3, 35.1) | 31.6 (22.4, 38.8) |

| Age at first ABCGRT initiation - median (IQR) | 31.3 (25.3, 38.8) | 30.7 (24.4, 38.3) | 30.8 (24.9, 38.1) | 31.4 (26.1, 39.2) | 33.6 (27.8, 40.9) | 27.2 (20.0, 35.6) | 35.4 (24.2, 44.3) |

| Baseline EDSS - median (IQR) | 2 (1, 2.5) | 2 (1, 2.5) | 1.5 (1, 2.2) | 2 (1, 2.5) | 2 (1, 2.5) | 2 (1, 2.5) | 1.5 (1, 2.5) |

| EDSS change on treatment - median (IQR) | 0 (0.5, 0.5) | 0 (0.5, 0.5) | 0 (0.5, 0.5) | 0 (0.5, 0.5) | 0 (0.5, 0.5) | 0 (0.5, 0.5) | 0 (0.5, 0.5) |

| ARR on treatment - mean (SD) | 0.53 (1.33) | 0.59 (1.38) | 0.47 (1.07) | 0.60 (1.40) | 0.58 (1.74) | 0.13 (0.36) | 0.23 (0.57) |

<table>
<thead>
<tr>
<th>Baseline MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1+ T1 gadolinium enhancing lesion</td>
</tr>
<tr>
<td>9+ T2 hyperintensive lesions</td>
</tr>
<tr>
<td>1+ infratentorial lesion</td>
</tr>
<tr>
<td>1+ juxtatentorial lesion</td>
</tr>
<tr>
<td>2+ periventricular lesions</td>
</tr>
</tbody>
</table>
Of these 2003 first DMT commencements, 615 (30.7%) initiated first IM-IFNβ-1a, 612 (30.6%) SC-IFNβ-1a, 380 (19.0%) IFNβ-1b, 305 (15.2%) glatiramer acetate, 54 (2.7%) natalizumab and 37 (1.8%) fingolimod. Median (IQR) time between CIS onset and treatment initiation across all first DMTs was 7.2 months (3.7, 13.3), ranging from a median 5.7 months in IM-IFNβ-1a first initiations to 17.1 months in the comparatively smaller fingolimod group. Median (IQR) age at treatment initiation was 31.3 years (25.3, 38.8) and were youngest in the natalizumab group (median 27.2 years) and oldest in fingolimod initiators (median 35.4 years). Mean (SD) time on treatment across all first DMTs was 2.3 years (2.0) with IM-IFNβ-1a recording the highest mean treatment duration (2.6 years, SD 2.1). Mean (SD) ARR on treatment was 0.53 (1.33) relapses per person-year of follow-up and was highest on IFNβ-1b (mean 0.6, SD 1.4) and lowest on natalizumab (mean 0.1, SD 0.4). Median (IQR) EDSS change across all first DMTs was 0 (-0.5, 0.5). There was no difference in EDSS change across any of the individual DMT groups (p=0.7811).

Of the 2003 first DMT initiations, 1076 (53.7%) ceased during follow-up at an incidence rate of 23.7 discontinuations per 100 person-years of follow-up (95% CI 22.3, 25.1). Mean (SD) time to discontinuation was 1.8 years (1.6). Discontinuation incidence ranged from 28.2 discontinuations per 100 person-years on IM-IFNβ-1a 28.2 (95% CI 25.42, 31.27) to 19.4 on natalizumab (95% CI 11.9, 31.6). The number of discontinuation events in the fingolimod group were too small to enable sufficiently reliable estimates of incidence.

Across all first DMTs, gender, country, DMT product identity, baseline EDSS and ARR were all associated with discontinuation on adjusted modelling (Table 2). Specifically females were significantly associated with 1.46 times the rate of discontinuation compared with males whilst patients managed within Italian, Canadian, Spanish, Czech and Turkish clinics were associated with 55%, 60%, 59%, 72% and 43% reductions in the rate of discontinuation compared with patients managed in Australian clinics. Every 1 unit increase in baseline EDSS was associated with 1.1 times
the rate of subsequent treatment discontinuation (HR 1.10, 95% CI 1.03, 1.18, p=0.003) whilst every 1 relapse per person-year increase in ARR was associated with 1.35 times the rate of discontinuation. By contrast, EDSS change on treatment was not associated with a significantly changed risk of discontinuation. Although older age at both CIS onset and first DMT start were associated with decreased risk of discontinuation, these associations fell out of significance on adjusted modelling. Furthermore, time from disease onset to commencement of first DMT was not predictive. None of the baseline cerebral MRI characteristics demonstrated or suggested association with discontinuation on either adjusted or unadjusted modelling.
### Table 2 – Predictors of treatment discontinuation – first DMD

<table>
<thead>
<tr>
<th>Practice</th>
<th>Gender</th>
<th>Date</th>
<th>Baseline characteristics</th>
<th>Baseline DMD characteristics</th>
<th>DDMC change (per follow-up period)</th>
<th>Length of DDMC</th>
<th>Clinic country</th>
<th>Time from onset to first DMT initiation (months)</th>
<th>HR (95% CI) p-value</th>
<th>HR (95% CI) p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age &lt; 18 months</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age 18-23 months</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age &gt; 23 months</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0</td>
<td>0</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
<td>Italy</td>
<td>2-2.5</td>
<td>1.00 (1.00, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
</tbody>
</table>

---

a. Hazard proportionality test: p=0.4519
b. Hazard proportionality test: p=0.3966
c. Hazard proportionality test: p=0.4212
d. Hazard proportionality test: p=0.3216
e. Hazard proportionality test: p=0.4091
Whilst all DMT products demonstrated a decreased rate of discontinuation compared with SC-IFNβ-1a, only IFNβ-1b (HR 0.70, 95% CI 0.55, 0.88, p=0.003) and natalizumab (HR 0.51, 95% CI 0.27, 0.95, p=0.034) remained predictive in the fully adjusted model (Figure 1). Similar clusters of associations was demonstrated when the modelling was run separately for each DMT, although this varied with the actual DMT product (Table 1). Female gender remained predictive of increased discontinuation rate but only for IM-IFNβ-1a and SC-IFNβ-1a. Similarly the combination of countries which remained predictive of decreased rates of discontinuation varied with DMT. Interestingly, increasing baseline EDSS remained associated with an increased risk of discontinuation when IM-IFNβ-1a was studied (HR 1.20, 95% CI 1.06, 1.36, p=0.005), however become very non-significant when the modelling was limited to SC-IFNβ-1a (HR 1.00, 95% CI 0.90, 1.11, p=0.932). Of all the a priori predictors studied, only on-treatment ARR remained consistently predictive of increase discontinuation rate across all DMT products where sufficient discontinuation events were available to permit fully adjusted modelling.

Figure 1 – Time to first DMD discontinuation by product identity
To investigate whether these associations varied by particular on-treatment time intervals, we repeated the analysis censoring the adjusted models at 6, 12, 18 and 24 months respectively (Table 3). Differences in discontinuation rate by DMT identity were most marked early on in a patient’s treatment with both the percent reductions in discontinuation rate and level of significance being greatest in first 6 months only model and gradually decreasing across the 12, 18 and 24 month models. A similar diminution of effect was observed with gender whilst country influences waxed and waned over time depending upon the specific country. Conversely the association of ARR and baseline EDSS remained consistent regardless of the censor point used.
Table 3 - Predictors of first DMD treatment discontinuation using different censor points

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Incidence per 100 person-years (95% CI):</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at onset (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at first DMT initiation (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline EDSS (categorical)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline EDSS (continuous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First DMT identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline BDI (continuous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incidence risk factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Mini-ITQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDSS change over follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow up: Discontinuations across full follow-up</td>
<td>Discontinuations - n (%): 258 (12.9) Discontinuations - n (%): 453 (22.6)</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- All first ABCGRT combined (n=2003)
- Predictors of first DMD treatment discontinuation using different censor points
- Incidence per 100 person-years (95% CI):
  - All first ABCGRT combined (n=2003)
- p-value:
  - Unadjusted
  - Adjusted
  - Omitted - colinear

**References:**
- Timothy Denis Spelman (58172)
As a further sensitivity analysis we reran the same set of predictors over models of any DMT discontinuation, not just limited to first DMTs (Table 4) using the time to multiple event modelling described in the methods. Of the original sample of 2003 patients, 828 (41.3%) reported two or more unique DMT commencements across their follow-up contributing a total of 3257 DMT initiations. Of these, 1662 (51.0%) ceased at an incidence rate of 25.1 discontinuations per 100 person-years of follow-up. With the exception of IFNβ-1b, all other DMTs were associated with decreased rates of any DMT discontinuation when compared with SC-IFNβ-1a, although the size of these reductions were substantially smaller when compared to the first DMT only discontinuation modelling. Similarly gender, on-treatment ARR and baseline EDSS remained predictive in this model although again the size of the Hazard Ratios were smaller when compared to the first DMT only model. First DMT product remained predictive which is not altogether unexpected given first DMTs account for over 50% of this “any DMT” sample. Interestingly older age at CIS onset, which dropped out of significance in the adjusted first DMT model after demonstrating significance on the unadjusted modelling, now remains predictive of a decreased rate of discontinuation when applied to the “any DMT” dataset.
Table 4 - Predictors of treatment discontinuation – any DMY and post-first DMT

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Incidence per 100 person-years (95% CI):</th>
<th>Incidence per 100 person-years (95% CI):</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-IFNβ-1a</td>
<td>25.10 (23.92, 26.34)</td>
<td>28.26 (26.06, 30.64)</td>
</tr>
<tr>
<td>IM-IFNβ-1a</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>IFNβ-1b</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.76 (0.67, 0.85)</td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Natalizumab</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Fingolimod</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>First DMT identity</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>SC-IFNβ-1a</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>IM-IFNβ-1a</td>
<td>0.87 (0.77, 0.97)</td>
<td>0.87 (0.77, 0.97)</td>
</tr>
<tr>
<td>IFNβ-1b</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Natalizumab</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Fingolimod</td>
<td>0.86 (0.75, 0.99)</td>
<td>0.86 (0.75, 0.99)</td>
</tr>
<tr>
<td>Gender</td>
<td>1.38 (1.22, 1.55)</td>
<td>1.38 (1.22, 1.55)</td>
</tr>
<tr>
<td>Male</td>
<td>1.38 (1.22, 1.55)</td>
<td>1.38 (1.22, 1.55)</td>
</tr>
<tr>
<td>Age at MS onset (5 years)</td>
<td>0.95 (0.92, 0.98)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>Age at first DMT initiation</td>
<td>0.95 (0.92, 0.98)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>ARR* on treatment</td>
<td>1.42 (1.35, 1.49)</td>
<td>1.42 (1.35, 1.49)</td>
</tr>
<tr>
<td>ARR* on treatment - stereoids</td>
<td>1.50 (1.44, 1.58)</td>
<td>1.50 (1.44, 1.58)</td>
</tr>
<tr>
<td>Clinic country</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Australia</td>
<td>0.69 (0.59, 0.80)</td>
<td>0.69 (0.59, 0.80)</td>
</tr>
<tr>
<td>Italy</td>
<td>0.54 (0.44, 0.65)</td>
<td>0.54 (0.44, 0.65)</td>
</tr>
<tr>
<td>Canada</td>
<td>0.54 (0.44, 0.65)</td>
<td>0.54 (0.44, 0.65)</td>
</tr>
<tr>
<td>Spain</td>
<td>0.54 (0.44, 0.65)</td>
<td>0.54 (0.44, 0.65)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.54 (0.44, 0.65)</td>
<td>0.54 (0.44, 0.65)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.54 (0.44, 0.65)</td>
<td>0.54 (0.44, 0.65)</td>
</tr>
<tr>
<td>Turkey</td>
<td>0.53 (0.39, 0.72)</td>
<td>0.53 (0.39, 0.72)</td>
</tr>
<tr>
<td>Other</td>
<td>0.53 (0.39, 0.72)</td>
<td>0.53 (0.39, 0.72)</td>
</tr>
<tr>
<td>Baseline EDSS (continuo)</td>
<td>1.05 (1.02, 1.09)</td>
<td>1.05 (1.02, 1.09)</td>
</tr>
<tr>
<td>Baseline EDSS (categoric)</td>
<td>1.05 (1.02, 1.09)</td>
<td>1.05 (1.02, 1.09)</td>
</tr>
<tr>
<td>1-1.5</td>
<td>0.82 (0.69, 0.96)</td>
<td>0.82 (0.69, 0.96)</td>
</tr>
<tr>
<td>2-2.5</td>
<td>0.82 (0.69, 0.96)</td>
<td>0.82 (0.69, 0.96)</td>
</tr>
<tr>
<td>3-3.5</td>
<td>0.82 (0.69, 0.96)</td>
<td>0.82 (0.69, 0.96)</td>
</tr>
<tr>
<td>4+</td>
<td>0.82 (0.69, 0.96)</td>
<td>0.82 (0.69, 0.96)</td>
</tr>
<tr>
<td>EDSS change (over follow-up)</td>
<td>1.08 (1.04, 1.12)</td>
<td>1.08 (1.04, 1.12)</td>
</tr>
<tr>
<td>Baseline MRI characteristics</td>
<td>1.08 (1.04, 1.12)</td>
<td>1.08 (1.04, 1.12)</td>
</tr>
<tr>
<td>McDonald T1 Gadolinium</td>
<td>0.98 (0.85, 1.13)</td>
<td>0.98 (0.85, 1.13)</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.08 (0.96, 1.22)</td>
<td>1.08 (0.96, 1.22)</td>
</tr>
<tr>
<td>McDonald T2 hyperinten</td>
<td>1.08 (0.96, 1.22)</td>
<td>1.08 (0.96, 1.22)</td>
</tr>
<tr>
<td>1-2</td>
<td>1.55 (0.97, 2.49)</td>
<td>1.55 (0.97, 2.49)</td>
</tr>
<tr>
<td>3-8</td>
<td>1.55 (0.97, 2.49)</td>
<td>1.55 (0.97, 2.49)</td>
</tr>
<tr>
<td>9+</td>
<td>1.55 (0.97, 2.49)</td>
<td>1.55 (0.97, 2.49)</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.55 (0.97, 2.49)</td>
<td>1.55 (0.97, 2.49)</td>
</tr>
<tr>
<td>McDonald Infratentorial</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1+</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>McDonald Juxtatentorial</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1-2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>McDonald Periventriculi</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1-2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Not reported</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Finally we reran the same multiple event modelling excluding the first DMT initiations (modelling post-first DMT discontinuation only) (Table 4). Although this reduced the number of DMT initiations to 1254, female gender, higher baseline EDSS and increasing on-treatment ARR remained predictive of an increased rate of discontinuation. Of the DMT product identities, only non-first natalizumab and fingolimod remained associated with a decreased risk of discontinuation. In contrast to the “first DMT only” modelling, every 1 unit increase in EDSS across follow-up was significantly associated with 1.11 and 1.17 times the rate of discontinuation in the “all DMTs” (HR 1.11, 95% CI 1.07, 1.15, p<0.001) and “all post-first DMTs” (HR 1.17, 95% CI 1.10, 1.25, p<0.001) respectively.

Discussion

Identification of patient and disease characteristics that are associated with higher probability of treatment discontinuation in clinical practice may enable the clinician to flag subsets of patients that may benefit from closer, more intensive follow-up whilst on treatment, facilitating more timely interventions and ultimately, more favourable patient outcomes. This large multinational observational study demonstrates that gender, clinic country, DMT identity, baseline EDSS and on-treatment ARR are all independently associated with the probability and timing of first DMT discontinuation. These results largely corroborate those suggested in a comparable, albeit smaller-scale analysis similarly based on the MSBASIS study with a couple of notable exceptions. First increasing on-treatment EDSS change, associated with an increased rate of discontinuation in the Meyniel study, no longer demonstrated a significant association in our larger study. This difference may in part be explained by the inclusion of baseline EDSS as a predictor in the discontinuation model, an explanatory variable omitted from the Meyniel model. Although no significant interaction between baseline EDSS and on-treatment EDSS change was demonstrated, our modelling suggests baseline EDSS is a better performer as an explanatory variable (i.e. explains more variation in the outcome) compared with EDSS change. From a clinical utility perspective, a metric such as baseline
EDSS, which is assessable soon before or at the point of first DMT initiation, has clear advantages as a predictor of subsequent treatment persistence compared with on-treatment EDSS change metric which, by definition, is only able to be derived at the point of treatment discontinuation – thus any predictive utility, at least at the level of the individual patient, is in hindsight only. Sensitivity analysis did suggest that EDSS change does again become predictive of discontinuation in subsequent (post-first) DMT courses which is not unexpected as the influence of first DMT baseline EDSS wanes with time.

A second important distinction from the Meyniel model is the pattern of association between the separate first DMT products and discontinuation rates. In the Meyniel study, of the four main DMT groups studied (injectable IFNs and glatiramer acetate), patients on first glatiramer acetate were associated with both the smallest incidence and risk of discontinuation, whilst SC-IFNβ-1a recorded the highest risk. By comparison our study observed the lowest incidence of first treatment discontinuation associated with IM-IFNβ-1a (22.4 discontinuations per 100 person-years) whilst the greatest risk reduction when modelled in a fully adjusted model was associated with IFNβ-1b when compared with SC-IFNβ-1a. These results are consistent with the large, retrospective United States cohort comparison of the three interferon-based agents and glatiramer acetate which observed greatest treatment adherence associated with IM-IFNβ-1a,speculating this may be linked to its less frequent dosing schedule. Of interest, our sensitivity analysis suggests that the identity of the first DMT product influences the probability of treatment persistence in subsequent DMT courses, underscoring the importance of the initial decision over which DMT to commence as a first agent and the importance of developing a replicable set of baseline or early predictors of persistence.

Natalizumab and fingolimod, not studied previously in this context, demonstrated even greater reductions in the rate of discontinuation when compared to SC-IFNβ-1a, however the comparatively small number of discontinuation events suggests that these estimates are likely underpowered and
thus would advise caution when making any inferences from these results. As both natalizumab and fingolimod become more common first-line, first-offered DMTs in more settings globally, it would be interesting to repeat this analysis and assess how our predictive model performs with greater follow-up on first natalizumab/fingolimod.

Trialling our model over a series of successively longer censor points suggests the relationship between some of our explanatory variables and the discontinuation outcome are not necessarily uniform or constant over time. The influence of both gender and DMT identity upon the probability of subsequent discontinuation were most pronounced within the first 6 months on first DMT. Conversely the probability of discontinuation by baseline EDSS and on-treatment ARR remained constant across this series of censor sets. In practice, these findings may assist the clinician, when considering commencing a patient on a first DMT, to anticipate which subsets of patients may experience difficulty with persistence, particularly within the first 6 to 12 months of therapy, and thus influence both the timing and choice of DMT product.

Regardless of the censoring points used or which DMT set modelled (first DMTs only, all DMT initiations, post-first DMT only) female gender, higher EDSS at baseline and increasing on-treatment ARR remained consistently predictive of increased rates of discontinuation and thus form the most reliable predictors of treatment discontinuation. Only the magnitude of the associations varied with generally larger effect sizes demonstrated in first DMT discontinuations compared with either the all DMT discontinuation or post-first DMT discontinuation modelling.

Consistent with previous studies, we observed discontinuation rates between different countries to vary markedly with patients managed in Australian clinics associated with significantly higher rates of discontinuation compared with patients managed in Italian, Canadian, Spanish, Czech or Turkish clinics. Although difficult to quantify, is has been speculated that part of the rationale underscoring
such location-specific discrepancies may include differences in health care delivery, the absence or presence of standardised treatment protocol and variance in the therapeutic relationship between patient and clinician.\textsuperscript{12,26-28} Importantly for our analyses the patient, disease and treatment characteristics described above remain predictive after controlling for country differences.

**Funding**

The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL.

**MSBasis study group co-investigators and contributors:** From the MS-Centrum Nijmegen, Nijmegen, The Netherlands, Dr Cees Zwanikken; from Flinders Medical Centre, Adelaide, Australia, Dr Mark Slee; from Geelong Hospital, Geelong, Australia, Dr Cameron Shaw; from the Brain and Medical Research Institute, University of Sydney, Sydney, Australia, Dr Michael Barnett; from St Vincent’s Hospital, Fitzroy, Australia, Dr Neil Shuey; from C. Mondino National Neurological Institute, Pavia, Italy Dr Roberto Bergamaschi; from Craigavon Area Hospital, Portadown, United Kingdom, Dr Stella Hughes; from INEBA, Buenos Aires, Argentina, Dr Maria Laura Saladino; from Mater Dei Hospital, Malta; Dr Norbert Vella; from Department NEUROFARBA, Section of Neurosciences, University of Florence, Florence, Italy, Dr Maria Pia Amato; from Hopital Tenon, Paris, France, Dr Etienne Roullet; from UT Southwestern, Dallas, United States, Dr Benjamin Greenberg; from Central Clinical Emergency Military Hospital, Bucharest, Romania, Dr Carmen-Adella Sirbu; from Hospital Angeles de las Lomas, Instituto Mexicano de Neurociencias, Huixquilucan de Degollado, Mexico, Dr Eli Skromne; from University Clinic of Neurology, Skopje, Macedonia, Dr Vladimir Bojkovski; from Universita Napoli, Napoli, Italy, Dr Pietro Carrieri; from The Walton Centre
Timothy Denis Spelman (58172)

for Neurology and Neurosurgery, Liverpool, United Kingdom, Dr Carolyn Young; from Hospital Donostia, Gipuzkoa, Spain, Dr Javier Olascoaga; from Consultorio Privado, Buenos Aires, Argentina, Dr Aldo Savino; and from Instituto de Neurociencias Cordoba, Cordoba, Argentina, Dr Elizabeth Alejandra Bacile. No compensation has been received for the persons who have made substantial contributions to the work but do not qualify as authors.

References


**G. ANALYSIS** - Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment settings and products

**G.1 Comparative efficacy of switching to natalizumab in active relapsing multiple sclerosis**

Title
Comparative efficacy of switching to natalizumab in active relapsing multiple sclerosis

Authors
T Spelman1*, T Kalincik1, A Zhang2, F. Pellegrini3, H Wiendi4, L Kappos5, L Tsvetkova6, S Belachew7, R Hyde8, F Verheul9, F Grand-Maison8, G Izquierdo9, P Grammond10, P Duquette11, A Lugaresi12, J Lechner-Scott13, C Oreja-Guevara14, R Hupperts15, T Petersen16, M Barnett17, M Trojano*3, H Butzkueven*1,18 on behalf of the MSBase Investigators and the TOP investigators.

1 Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia
2 Biogen Idec Inc., Cambridge, MA
3 Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy
4 Department of Neurology, University of Münster, Münster, Germany
5 Department of Neurology, University Hospital Basel, Basel, Switzerland
6 Biogen Idec Int BV, Badhoevedorp, The Netherlands
7 Groene Hart Ziekenhuis, Gouda, The Netherlands
8 Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
9 Hospital Universitario Virgen Macarena, Sevilla, Spain
10 Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada
11 Hôpital Notre Dame, Montreal, Canada
12 MS Center, Department of Neuroscience, Imaging and Clinical Sciences, University ‘G. d’Annunzio’, Chieti, Italy
13 John Hunter Hospital, Newcastle, Australia
14 University Hospital San Carlos, Madrid, Spain
15 Orbis Medical Centre, Sittard-Geleen, The Netherlands
16 Kommunehospitalet, Arhus C, Denmark
17 Brain and Mind Research Institute, Sydney, Australia
18 Dept of Neurology, Eastern Health, Monash University, Box Hill, Australia

Contact information
Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia
tim@burnet.edu.au
Ph: 61 3 9342 4406
Fx: 61 3 9349 5997

Table and figure count limit
3 Tables, 1 supplementary table, 5 Figures, 1 supplementary figure
ABSTRACT

Objective: To compare treatment efficacy and persistence in patients who switched to natalizumab versus those who switched between glatiramer acetate (GA) and interferon-beta (IFNβ) after an on-treatment relapse on IFNβ or GA using propensity score matched real-world datasets.

Methods: Patients included were registered in MSBase or the TYSABRI Observational Program (TOP), had relapsed on IFNβ or GA within 12 months prior to switching to another therapy, and had initiated natalizumab or IFNβ/GA treatment ≤6 months after discontinuing prior therapy. Covariates were balanced across post-switch treatment groups by propensity score matching at treatment initiation. Relapse, persistence, and disability measures were compared between matched treatment arms in the total population (n=869/group) and in subgroups defined by prior treatment history (IFNβ only [n=578/group], GA only [n=165/group], or both IFNβ and GA [n=176/group]).

Results: Compared to switching between IFNβ and GA, switching to natalizumab reduced annualized relapse rate in year one by 65-75%, the risk of first relapse by 53-82% (mean follow-up 1.7-2.2 years) and treatment discontinuation events by 48-65% (all p≤0.001). In the total population, switching to natalizumab reduced the risk of confirmed disability progression by 26% (p=0.036) and decreased the total disability burden by 1.54 EDSS-years [p<0.0001] over the first 24 months post-switch.

Interpretation: Using large, real-world, propensity-matched datasets we demonstrate that after a relapse on IFNβ or GA, switching to natalizumab (rather than between IFNβ and GA) led to superior outcomes for patients in all measures assessed. Results were consistent regardless of the prior treatment identity.
INTRODUCTION

For patients with relapse on first-line interferon-beta (IFNβ) or glatiramer acetate (GA) therapy, switching to another immunomodulatory therapy is a potentially useful treatment strategy.\textsuperscript{1-4} There is some evidence that switching between IFNβ and GA and among different IFNβ therapies can improve patient treatment response.\textsuperscript{5-7} On the other hand, natalizumab (TYSABRI\textsuperscript{®}) is also recommended for patients with relapse on IFNβ and/or GA therapy.\textsuperscript{8,9} Improvements in disability status and ambulation have been reported in single-arm observational studies of patients who switched to natalizumab after experiencing high disease activity while on another disease-modifying therapy (DMT), usually IFNβ/GA.\textsuperscript{10,11} In a single centre retrospective analysis by Rio and colleagues,\textsuperscript{6} relapse rates in a IFNβ/GA treatment failure population declined significantly after switch to another IFN/GA product or switch to natalizumab, but these two switch groups were not compared. Another two-centre 24-month observational study showed that patients who switched to natalizumab (n=106) or IFNβ/GA (n=161) after first-line (IFNβ/GA) treatment failure were more likely to remain free of relapse, disability progression, and MRI disease activity than those who switched between IFNβ and GA formulations,\textsuperscript{12} with all treatment effects non-significant in year one but significant in year 2.

These treatment comparisons are challenging to interpret because treatment assignations are non-random; and bias may be introduced due to differing patient characteristics. Typically, patients switching to natalizumab would be expected to have more severe disease than patients switching between IFNβ or GA treatments, with resultant bias against natalizumab in outcome analyses. Given the limited observational evidence and lack of randomised controlled trial evidence, we examined treatment outcomes of switch to natalizumab versus switch between IFNβ and GA therapy after IFNβ/GA failure using patients (n=869/group) matched by disease severity and demographic variables at the time of switch.
The aims of this study were to compare relapse rate, treatment persistence, and disability progression in MS patients who switched therapy after failure on Betaferon®, Betaseron®, Rebif®, Avonex®, Copaxone® or Extavia® (BRACE) treatments using propensity matched samples from the MSBase Observational Registry and the TYSABRI Observational Program (TOP), two distinct real-world cohorts with contemporaneous recruitment, according to prospectively defined protocols. We decided to apply propensity score matching, a powerful statistical technique for correcting multiple baseline covariate imbalances in non-randomly selected cohorts\textsuperscript{13-15}. Recently this technique was successfully used to aid comparisons of IFNβ treatment persistence and disease outcomes using observational data from the MSBase registry.\textsuperscript{16}

MATERIALS & METHODS

Patient sources

MSBase Registry

Patients in the BRACE treatment arms of this study were sourced from the international online MSBase Registry. The MSBase Registry was established in 2004 to collect disease-related information from consenting patients attending MS clinics. As of April 4\textsuperscript{th}, 2013, a total of 21,348 people with MS across 60 clinics in 26 countries were participating in MSBase. The registry’s member centres, almost exclusively large academic MS centres, follow a defined minimum dataset protocol to prospectively collate outcomes data using an internet-based, physician owned and operated system [www.msbase.org].\textsuperscript{17} Each center enters patient data either in the offline iMed\textsuperscript{©} local electronic database or the online MSBase registry data entry system during routine clinic visits and intermittently uploads codified datasets to the MSBase server. Physicians record clinical information such as date of MS onset, diagnostic category, Kurtzke Expanded Disability Status Scale (EDSS) score, relapse onset dates and characteristics, cerebral MRI, and other investigations, and commit to minimum annual follow-up. A clinical attack is defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 hours, in the absence of concurrent
illness or fever, and occurring at least 30 days after a previous attack, also previously applied in an MSBase relapse phenotype analysis.\textsuperscript{18,19} Records are classified as complete and eligible for analyses if they meet a minimum required dataset. Quality of EDSS assessment is monitored by Neurostatus certification of investigators. Informed consent (as required by local laws and regulations) is provided by each participant in MSBase. At each contributing center the project has Human Research Ethics Committee approval or exemption.

\textit{TYSABRI Observational Program (TOP)}

Patients treated with natalizumab were sourced from TOP (Biogen Idec, Cambridge, MA USA). TOP is an ongoing, open-label, multinational, multicentre, prospective, observational study conducted in clinical practice settings in Europe, Australia, Canada and Argentina.\textsuperscript{20} Patients are recruited within 3 months of commencing natalizumab. Data are collected at regular clinical visits every 6 months. Data entry is web-based. The primary endpoint is long-term safety (incidence and type of serious adverse events). Secondary endpoints include measures of MS disease activity (including the occurrence of clinical relapses and change in EDSS score). In TOP, a clinical relapse is also defined as new or recurrent neurological symptoms, not associated with fever, lasting for $\geq 24$ hours and followed by a period of 30 days of stability or improvement. Study endpoints are assessed uniformly across sites. To assure standardised examinations and consistent definitions for the EDSS Functional System (FS) scores, participating physicians are provided a copy of the interactive Neurostatus Training DVD-ROM, and Neurostatus certification is highly recommended. Investigators not previously certified are offered the same online certification (http://www.neurostatus.net) available to MSBase investigators. To reduce the risk of entry error with EDSS score reporting, electronic case report forms (CRFs) were designed to automatically generate queries for data inconsistencies, including data that were out of range or otherwise invalid. The CRF calculates an EDSS score based on the Kurtzke FS and ambulation scores that were entered. The same EDSS calculator was used to assist MSBase investigators. Data quality control procedures checked for consistency across data sets.
Site-based verification and correction was used for residual data queries. At the time of extraction, there were 4,821 patients participating in TOP across 16 countries. The TOP study design is in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines, and all enrolled patients provided written informed consent.

**Study design**

*Patients and Subgroups*

Participants in MSBase or TOP who relapsed in the 12 months prior to switching to natalizumab or a BRACE therapy and who initiated the new treatment ≤6 months after discontinuation of the prior BRACE therapy were included in this study. Treatment efficacy was initially compared between patients who switched from any BRACE therapy to natalizumab and those who switched between different BRACE therapies. Additional comparisons were performed using 3 subgroups of patients (based on prior BRACE treatment): 1) those with only prior IFNβ therapy exposure who switched to either natalizumab or GA; 2) those with only prior GA therapy exposure who switched to natalizumab or IFNβ; and 3) those with both prior GA and IFNβ therapy exposure who switched to natalizumab or another IFNβ therapy. Propensity score matching was performed separately for the total patient population and each subgroup (Figure 1).
**Figure 1: Study profile**

**Efficacy Measures**

The primary efficacy outcomes assessed were annualized relapse rate (ARR), time to first relapse on therapy, time to treatment discontinuation and time to confirmed disability progression. An area under the disability/time curve (AUC) analysis was conducted as a secondary, exploratory outcome.

Confirmed disability progression events were defined as ≥3-month confirmed increases of ≥0.5 points for patients with a baseline EDSS score >5.5, ≥1.0 point for those with a baseline EDSS score between 1.0 and 5.5, inclusive, and ≥1.5 points for those with a baseline EDSS score of 0. EDSS scores recorded within 30 days after the onset of a relapse were excluded. A minimum of 3 visits (including baseline) at which an EDSS was formally recorded were, by definition, required to be able to assess confirmed disability progression. Thus, this analysis was limited to patients with a minimum of 3 EDSS scores reported. Since this decreased the number of matched pairs available, comparisons of the time to confirmed disability progression were not performed across treatment arms in the subgroup analyses.
As an exploratory analysis, AUC comparisons were performed from baseline through 24 months of treatment. AUC values were derived for both the natalizumab and BRACE treatment arms to estimate each patient’s total study experience with respect to disability burden after switching treatment. Only propensity-matched patients who remained on the new therapy ≥24 months after switching and who had an EDSS score recorded within ±6 months of the 24-month post-baseline mark were included in this analysis. This limited the number of matched pairs available. Therefore, AUC measures were not compared in the subgroup analyses.

Statistical analyses

The data from both registries were combined according to a pre-specified protocol. Categorical variables were summarized using frequency and percentage. Continuous variables were assessed for significant departures from normality using a Shapiro-Wilk test of skew and summarised using mean and standard deviation (SD) or standard error (SE), or median and inter-quartile range (IQR) as appropriate. All baseline covariates common to and available from both registries were used to calculate propensity scores. These included sex, age, disease duration, baseline EDSS, number of DMT initiations, duration of DMT use as a proportion of disease duration, total relapses and total steroid-treated relapses in the 12 and 24 months preceding baseline. Prior DMTs reported were Betaferon®, Betaseron®, Rebif®, Avonex®, Copaxone®, or Extavia®.

Propensity scores were calculated for each individual patient and represented the probability that a patient from either registry would have commenced natalizumab treatment based purely on pre-treatment baseline characteristics. This propensity score was derived from a logistic regression model, in which receipt of natalizumab was the outcome variable and the pre-treatment characteristics formed the explanatory variables. Patients from each treatment arm were matched, based on high similarity of propensity score, on a 1:1 basis using a 5-to-1 digit matching algorithm with a 0.01 calliper. Success of matching was assessed using both paired tests and analysis of
standardized differences. Imbalance was defined as an absolute value of the standardized difference equal or greater than 0.20. Wilcoxon rank-sum and chi-square tests were used to compare unmatched baseline characteristics by treatment arm as appropriate. Wilcoxon signed-rank and McNemar tests were used to compare baseline characteristics in the matched data for continuous variables and proportions, respectively. Standardized differences were calculated for both unmatched and matched comparisons, permitting direct comparison of different baseline characteristics with the same standardized units. Baseline for these analyses was defined as the date the patient initiated the new treatment after BRACE therapy discontinuation. Post-hoc Rosenbaum sensitivity analyses across all outcomes were conducted to test the sensitivity of our propensity-matched models to unobserved heterogeneity.

Comparative analyses of time to first relapse on treatment, time to treatment discontinuation, and time to confirmed disability progression between natalizumab and BRACE comparator groups were performed using a Cox Marginal Model, clustered for the matched pair. Hazard proportionality was assessed via analysis of scaled Schoenfeld residuals and for all models presented in this report, hazard proportionality was satisfied. A competing risks extension of the Cox time-to-event model was used to further assess for the influence of informative censoring on event ascertainment, for example the influence of differences in follow-up time (censoring at last observed assessment) by treatment arm on observing relapse or discontinuation events.

In order to perform AUC analyses, for each patient all EDSS scores recorded within the 24-month interval were plotted, and the inter-EDSS serial disability area was calculated as the product of the difference between observed and baseline EDSS and the time elapsed between two successive EDSS assessments (Figure 2). This quantity could take positive values (for disability area recorded above baseline EDSS) or negative values (area contributed by EDSS recorded below baseline EDSS). These inter-EDSS AUC values were then summed to produce a cumulative AUC across the 24-month on-
treatment interval. This approach to calculating AUC presumes EDSS change between two successive assessments is not a constant, linear change as has been assumed in comparable studies\textsuperscript{30-32} but rather an event-based step up or down in EDSS, more consistent with the attack/relapse course of MS. As a sensitivity analysis we compared the performance of our AUC approach with that calculated using the trapezoidal rule,\textsuperscript{33,34} which presumes a steady, linear change in EDSS between assessment points. The standardised 24-month AUC values were compared across propensity score matched treatment arms with quantile median regression using Censored Least Absolute Deviations (CLAD)\textsuperscript{35,36} to adjust for the matched pairs. A quantile regression of the median was preferred to simple linear regression of the mean because the distribution of standardised 24-month AUC values was significantly skewed and resistant to transformation. A Cochran-Armitage test was used to check for non-linearity in the associations between AUC and treatment arm. Data was extracted and compiled on April 4\textsuperscript{th}, 2013. To test for potential ascertainment bias secondary to unequal frequency of EDSS assessment points and/or unequal time between assessments by switch therapy arm, we conducted a sensitivity analysis limiting the serial disability-time analysis to just that subset of patients who recorded exactly 5 assessments (at baseline, 6, 12, 18 and 24 months). Consistent with the real-world nature of the data, temporal assessments were not always made exactly on the 6-month mark, thus we analysed a series of time ranges around each of the 6, 12, 18 and 24 month points – namely 1, 1.5 and 2 months. Informative censoring was minimized through application of pairwise censoring. This involved censoring both members of the propensity-matched natalizumab-\textit{BRACE} pair at the earliest censor point recorded by either member of the pair. All analyses were undertaken using Stata version 12 (StataCorp, College Station, Texas) and R (R Foundation for Statistical Computing, Vienna, Austria).
Figure 2: Example of a cumulative AUC measurement from a sample 24-month EDSS/time plot

The red line represents baseline EDSS. Dots indicate individual EDSS measurements. The area above baseline EDSS (blue) minus the area below baseline EDSS (purple) equals the cumulative summed AUC.

RESULTS

Patients with relapse on any BRACE therapy who switched to another BRACE therapy or to natalizumab

Patients

Of the 3377 patients in TOP who met inclusion criteria and switched from BRACE to natalizumab and the 1147 patients in MSBase who met inclusion criteria and switched between BRACE treatments, 869 were successfully paired by propensity score. Distribution plots of propensity scores by switch arm for both the unmatched and matched sample are presented in supplemental Figure 1 for the primary any BRACE switch analysis group. As expected, baseline covariates were markedly different between treatment groups in the unmatched sample, with greater disease severity demonstrated in
the patients switching to natalizumab (Table 1); no significant differences were demonstrated after propensity score matching (Table 2). For the propensity score matched patients, mean (SD) follow-up from baseline was 2.24 years (2.47) in the BRACE group compared to 1.95 years (1.23) in the natalizumab group ($p = 0.002$). Mean (SD) time between on-treatment assessments was 6.33 (3.91) months in the BRACE group and 6.30 (1.83) months in the natalizumab group ($p = 0.753$).
### Table 1: Pre-matching comparison of baseline characteristics by switch group

<table>
<thead>
<tr>
<th>Variable (at baseline)</th>
<th>Total Population</th>
<th>Prior therapy with IFNβ</th>
<th>Prior therapy with GA</th>
<th>Prior therapy with IFNβ and GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NTZ</td>
<td>IFNβ/GA</td>
<td>NTZ</td>
<td>GA</td>
</tr>
<tr>
<td>n</td>
<td>3377</td>
<td>1147</td>
<td>2109</td>
<td>675</td>
</tr>
<tr>
<td>Female, % p-value</td>
<td>72.0 (0.001)</td>
<td>&lt;0.001</td>
<td>70.4 (0.001)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>17.5</td>
<td>-16.5</td>
<td>14.6</td>
<td>-14.1</td>
</tr>
<tr>
<td>Age, median (IQR) p-value</td>
<td>37 (30,44)</td>
<td>37 (31,45)</td>
<td>37 (29,44)</td>
<td>37 (30,44)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-7.3</td>
<td>-8.9</td>
<td>0.048</td>
<td>0.064</td>
</tr>
<tr>
<td>Disease duration, median (IQR) p-value</td>
<td>6.9 (3.4,13.9)</td>
<td>6.4 (3.2,11.6)</td>
<td>6.3 (3.1,16)</td>
<td>5.8 (2.8,11)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.104</td>
<td>0.114</td>
<td>0.079</td>
<td>0.114</td>
</tr>
<tr>
<td>Proportion disease duration on DMT, median (IQR) p-value</td>
<td>0.5 (0.0,8)</td>
<td>0.4 (0.2,0.6)</td>
<td>0.6 (0.3,0.8)</td>
<td>0.4 (0.2,0.6)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>4.6</td>
<td>3.2</td>
<td>0.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Number of DMT starts mean (SD) p-value</td>
<td>1.5 (0.8)</td>
<td>1.4 (0.7)</td>
<td>1.3 (0.6)</td>
<td>1.4 (0.7)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.743</td>
<td>0.4</td>
<td>0.0016</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Number of DMT starts/ disease duration, mean (SD) p-value</td>
<td>0.5 (0.5)</td>
<td>0.4 (0.4)</td>
<td>0.3 (0.4)</td>
<td>0.4 (0.5)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.094</td>
<td>-0.159</td>
<td>0.0057</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Baseline ESSS, median (IQR) p-value</td>
<td>3 (2,4.5)</td>
<td>2.5 (2,4)</td>
<td>3 (2,4)</td>
<td>2.5 (2,4)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>45.6</td>
<td>47.3</td>
<td>4.0</td>
<td>4.3</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months, mean (SD) p-value</td>
<td>2.0 (1.0)</td>
<td>1.6 (0.9)</td>
<td>2.0 (1.0)</td>
<td>1.6 (0.9)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total steroid-treated relapses last 12 months, mean (SD) p-value</td>
<td>1.7 (1.1)</td>
<td>0.9 (0.9)</td>
<td>1.7 (1.1)</td>
<td>0.9 (0.9)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.707</td>
<td>0.707</td>
<td>0.707</td>
<td>0.707</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months, mean (SD) p-value</td>
<td>2.0 (1.1)</td>
<td>2.4 (1.4)</td>
<td>2.0 (1.1)</td>
<td>2.4 (1.4)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.914</td>
<td>0.914</td>
<td>0.914</td>
<td>0.914</td>
</tr>
<tr>
<td>Total steroid-treated relapses last 24 months, mean (SD) p-value</td>
<td>2.3 (1.5)</td>
<td>1.3 (1.3)</td>
<td>2.3 (1.5)</td>
<td>1.3 (1.3)</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

NTZ = natalizumab; GA = glatiramer acetate; IQR = interquartile range; SD = standard deviation

Legend for Table 1:

a. Imbalance defined as an absolute value ≥20%.
## Table 2: Propensity-matching comparison of baseline characteristics by switch group

<table>
<thead>
<tr>
<th>Variable (at baseline)</th>
<th>Total Population</th>
<th>Prior therapy with IFNβ/GA</th>
<th>Prior therapy with GA</th>
<th>Prior therapy with IFNβ and GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NTZ</td>
<td>IFNβ/GA</td>
<td>NTZ</td>
<td>GA</td>
</tr>
<tr>
<td>n</td>
<td>869</td>
<td>869</td>
<td>578</td>
<td>578</td>
</tr>
<tr>
<td>Female, %</td>
<td>77.2</td>
<td>78.9</td>
<td>76.5</td>
<td>76.1</td>
</tr>
<tr>
<td>p-value</td>
<td>0.412</td>
<td>0.412</td>
<td>0.9655</td>
<td>0.9655</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-4.2</td>
<td>-4.2</td>
<td>6.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>38</td>
<td>(31, 45)</td>
<td>37</td>
<td>(31, 45)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.518</td>
<td>0.518</td>
<td>0.8507</td>
<td>0.8507</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>3.2</td>
<td>3.2</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Disease duration, median (IQR)</td>
<td>6.8</td>
<td>(3.4, 13.0)</td>
<td>6.2</td>
<td>(3.0, 11.6)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.340</td>
<td>0.340</td>
<td>0.179</td>
<td>0.179</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>2.0</td>
<td>2.0</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Proportion disease duration on DMT, median (IQR)</td>
<td>0.4</td>
<td>(0.2, 0.6)</td>
<td>0.4</td>
<td>(0.2, 0.6)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.434</td>
<td>0.434</td>
<td>0.5044</td>
<td>0.5044</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-4.1</td>
<td>-4.1</td>
<td>-5.0</td>
<td>-5.0</td>
</tr>
<tr>
<td>Number of DMT starts mean (SD)</td>
<td>1.4</td>
<td>(0.7)</td>
<td>1.4</td>
<td>(0.7)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.178</td>
<td>0.178</td>
<td>0.6431</td>
<td>0.6431</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>3.9</td>
<td>3.9</td>
<td>-9.8</td>
<td>-9.8</td>
</tr>
<tr>
<td>Number of DMT starts / disease duration, mean (SD)</td>
<td>0.4</td>
<td>(0.6)</td>
<td>0.4</td>
<td>(0.4)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.533</td>
<td>0.533</td>
<td>0.1832</td>
<td>0.1832</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>1.1</td>
<td>1.1</td>
<td>-10.6</td>
<td>-10.6</td>
</tr>
<tr>
<td>Baseline EDSS, median (IQR)</td>
<td>3</td>
<td>(2, 4)</td>
<td>3</td>
<td>(2, 4)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.138</td>
<td>0.138</td>
<td>0.6519</td>
<td>0.6519</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>6.6</td>
<td>6.6</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months, mean (SD)</td>
<td>1.6</td>
<td>(0.7)</td>
<td>1.6</td>
<td>(0.9)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.670</td>
<td>0.670</td>
<td>0.7929</td>
<td>0.7929</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-18.8</td>
<td>-18.8</td>
</tr>
<tr>
<td>Total steroid-treated relapses last 12 months, mean (SD)</td>
<td>0.9</td>
<td>(0.8)</td>
<td>1.0</td>
<td>(0.9)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.090</td>
<td>0.090</td>
<td>0.2891</td>
<td>0.2891</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-6.2</td>
<td>-6.2</td>
<td>-8.4</td>
<td>-8.4</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months, mean (SD)</td>
<td>2.4</td>
<td>(1, 2)</td>
<td>2.4</td>
<td>(1, 2)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.407</td>
<td>0.407</td>
<td>0.9529</td>
<td>0.9529</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>2.3</td>
<td>2.3</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Total steroid-treated relapses last 24 months, mean (SD)</td>
<td>1.3</td>
<td>(1, 1)</td>
<td>1.4</td>
<td>(1, 2)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.198</td>
<td>0.198</td>
<td>0.5144</td>
<td>0.5144</td>
</tr>
<tr>
<td>Standardized difference, %</td>
<td>-4.3</td>
<td>-4.3</td>
<td>-1.5</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

NTZ—natalizumab; GA—glatiramer acetate; IQR—interquartile range; SD—standard deviation

Legend for Table 2

a. Imbalance defined as an absolute value ≥20%.
Relapse rate

During the first 12 months after treatment switch, ARR was higher in patients who switched to another BRACE therapy (mean, 0.58; SD, 0.86) than in those who switched to natalizumab (mean, 0.20; SD, 0.52) \( (p < 0.0001) \), representing a 66% relative reduction in ARR for patients who switched to natalizumab. This difference was sustained over subsequent years (Table 3). Over the study period, patients who switched to natalizumab had a 54% reduction in the risk of first relapse (hazard ratio [HR], 0.46; 95% confidence interval [CI], 0.39-0.53; \( p < 0.001 \)) (Figure 3A). When the analysis was limited to the first 12 months after treatment switch, the risk of first on-treatment relapse was reduced by 65% for patients who switched to natalizumab (HR, 0.35; 95% CI, 0.28-0.44; \( p < 0.001 \)).
### Table 3: Annualised relapse rates by treatment group and post-baseline year

<table>
<thead>
<tr>
<th>Year</th>
<th>Total population</th>
<th>Prior therapy with IFNβ</th>
<th>Prior therapy with GA</th>
<th>Prior therapy with IFNβ and GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of patients</td>
<td>ARR, mean (SD)</td>
<td>No. of patients</td>
<td>ARR, mean (SD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>Natalizumab</td>
<td>607 0.30 (0.53)</td>
<td>439 0.16 (0.50)</td>
<td>112 0.21 (0.47)</td>
</tr>
<tr>
<td></td>
<td>BRACE</td>
<td>497 0.58 (0.86)</td>
<td>369 0.54 (0.86)</td>
<td>113 0.60 (0.82)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Year 2</td>
<td>Natalizumab</td>
<td>372 0.18 (0.38)</td>
<td>269 0.19 (0.50)</td>
<td>63 0.24 (0.50)</td>
</tr>
<tr>
<td></td>
<td>BRACE</td>
<td>333 0.48 (0.59)</td>
<td>257 0.43 (0.75)</td>
<td>95 0.38 (0.70)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.1721</td>
</tr>
<tr>
<td>Year 3</td>
<td>Natalizumab</td>
<td>189 0.16 (0.28)</td>
<td>143 0.09 (0.18)</td>
<td>31 0.03 (0.18)</td>
</tr>
<tr>
<td></td>
<td>BRACE</td>
<td>219 0.39 (0.46)</td>
<td>167 0.29 (0.56)</td>
<td>81 0.10 (0.32)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.2513</td>
</tr>
<tr>
<td>Year 4</td>
<td>Natalizumab</td>
<td>59 0.14 (0.25)</td>
<td>62 0.00 (0.00)</td>
<td>10 0.10 (0.32)</td>
</tr>
<tr>
<td></td>
<td>BRACE</td>
<td>166 0.36 (0.42)</td>
<td>133 0.21 (0.49)</td>
<td>78 0.22 (0.47)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.0002</td>
<td>0.0009</td>
<td>0.4361</td>
</tr>
</tbody>
</table>

ARR=annualised relapse rate; SD=standard deviation

Legend for Table 3

*Patients were only included in the analysis for a given year if they were both on treatment and followed-up through the entire the year.
Figure 3: Time to A) first relapse, B) treatment discontinuation, or C) 3-month confirmed disability progression after treatment switch.

*Reference group switched to BRACE
†Reference group switched to natalizumab
Treatment persistence

For patients who switched to natalizumab there was a 60% reduction in the risk of further treatment discontinuation (HR, 0.40; 95% CI, 0.34-0.47; p < 0.001) (Figure 3B). Limiting the model to the first 12 months after treatment switch increased the effect size. Patients who switched to natalizumab had a 74% reduction in the risk of treatment discontinuation (HR, 0.26; 95% CI, 0.20-0.34, p < 0.001). Despite longer mean follow-up in the BRACE treatment arm, competing risks models for both the time to first relapse and treatment discontinuation confirmed that this difference did not significantly influence the estimated hazard ratios. Forty-one (4.7%) of the 869 matched patients who switched to natalizumab tested positive for anti-JCV antibodies at least once during the follow-up period.

Disability progression

Of the 869 propensity score matched pairs, a total of 374 patients in the BRACE arm and 514 patients in the natalizumab arm met the minimum requirement of 3 reported EDSS scores (including baseline) and were included in the disability progression analysis. There were no significant differences in baseline characteristics between the groups. Patients who switched to natalizumab had a 26% reduction in the risk of 3-month confirmed disability progression (HR, 0.74; 95% CI, 0.55-0.97; p = 0.036) (Figure 3C). There was no significant difference in time to confirmed disability progression when this analysis was limited to the first 12 months after treatment switch (HR, 0.86; 95% CI, 0.51-1.44; p = 0.561).

Of the 1738 propensity score matched patients in the total study population, 568 had the minimum on-treatment follow-up of 24 months required for inclusion in the AUC analysis. There were no significant differences in any of the baseline characteristics between the treatment groups, however ascertainment of EDSS during the 24-month interval was significantly more frequent in the BRACE
arm compared with the natalizumab arm. Within the 24-month AUC analysis time window, the mean (SD) number of visits with an EDSS score reported was 4.75 (1.72) in the BRACE arm and 3.67 (0.65) in the natalizumab arm ($p = 0.0001$).

Patients switching from BRACE to natalizumab had significantly less total disability burden, as measured by standardized 24-month AUC values, compared to patients who switched between BRACE therapies. The mean cumulative AUC was decreased by 0.64 EDSS-years in the natalizumab treatment arm compared to the BRACE treatment arm (natalizumab: mean [SD], -3.30 [1.65] EDSS-years; BRACE: -2.66 [1.70] EDSS-years; $p < 0.0001$). On quantile regression, median standardized 24-month AUC was decreased by 1.54 EDSS-years (95% CI -2.30, -0.78) ($p < 0.0001$) in patients who switched to natalizumab compared to those who switched to BRACE treatments. In a sensitivity analysis, AUC was measured using the trapezoidal rule$^{25-27}$; median standardized 24-month AUC was decreased by 1.01 EDSS-years in the natalizumab group compared to the BRACE group ($p = 0.001$) when this method was employed. For comparison, time to first 3-month confirmed disability progression was also assessed over the same 24-month time window and in the subset of patients used in the AUC analysis (including only propensity-matched patients who remained on the new therapy ≥24 months after switching). In this group, patients who switched to natalizumab had a 45% reduction in the risk of 3-month confirmed disability progression (HR, 0.55; 95% CI, 0.35-0.88; $p = 0.012$). To test this result for potential ascertainment bias secondary to unequal frequency of EDSS assessment points and/or unequal time between assessments by treatment arm, we reran this model limiting the analysis to just the subset of patients who recorded exactly five visits (at baseline, 6, 12, 18 and 24 months) within the 24-month consideration period. Applying, successively, a 1, 1.5 and 2-month time range buffer around each 6-month assessment point we observed a significant decrease in median standardized 24-month AUC of 1.68 EDSS-years (95% CI -2.51, -0.69), 1.59 EDSS-years (95% CI -2.47, -0.72) and 1.56 EDSS-years (95% CI -2.35, -0.75) respectively in the natalizumab switch arm relative to the BRACE switchers. These results are consistent with the primary analysis.
with only a marginal broadening of the confidence intervals around the point estimates consistent with the loss of sample associated with each sensitivity analyses. This suggests the results of the primary analysis are relatively resistant to any ascertainment bias conferred by the minor imbalances in 24-month EDSS assessment frequency and time between assessments.

**Subgroup analyses by prior BRACE treatment**

After propensity scores were used to match patients in TOP and MSBase independently in each prior treatment subgroup, there were no significant differences in the baseline characteristics between treatment arms in any subgroup (Table 2). Mean (SD) follow-up in subgroup 1 (patients with only prior IFNβ therapy who switched to GA or natalizumab) was 2.24 (2.30) years in the GA group compared with 1.98 (1.23) years in the natalizumab group; in subgroup 2 (patients with only prior GA therapy who switched to IFNβ or natalizumab) it was 1.69 (2.01) years in the IFNβ group compared with 1.73 (1.17) years in the natalizumab group; in subgroup 3 (patients with both prior IFNβ and GA therapy who switched to another IFNβ or natalizumab) it was 1.82 (1.94) years in the IFNβ group compared with 1.79 (1.13) years in the natalizumab group.

In all subgroups, on-treatment ARR in the first 12 months after switching treatment was lower in patients who switched to natalizumab compared to those who switched to another BRACE therapy; the relative reduction in ARR was 70% in subgroup 1, 65% in subgroup 2, and 75% in subgroup 3. ARR remained lower in the natalizumab group in subsequent years in all subgroups (Table 3). Patients who switched to natalizumab also had a lower risk of first relapse and a lower risk of treatment discontinuation compared to patients who switched to another BRACE therapy in all subgroups. Figures 4 and 5 display the Kaplan-Meier curves of time to first relapse and treatment discontinuation for each prior treatment subgroup; relative risk reductions and rate increases associated with each treatment decision are included in the text on each figure.
Figure 4: Time to first relapse after treatment switch by prior treatment subgroup, (A) IFNβ, (B) GA, or (C) IFNβ and GA
Figure 5: Time to treatment discontinuation after treatment switch by prior treatment subgroup, (A) IFNβ, (B) GA, or (C) IFNβ and GA

*Reference group switched to BRACE
†Reference group switched to natalizumab
Sensitivity analyses

To assess the sensitivity of our propensity matched models to potential confounding secondary to imbalance of baseline MRI metrics, we remodelled treatment arm as a predictor of first on-treatment relapse, switch therapy discontinuation and three-month confirmed disability progression incorporating baseline cerebral MRI metrics, where available, as adjusting covariates in the Cox marginal model. The reduction in the rate of first on-treatment relapse associated with natalizumab relative to BRACE observed in the matched primary analysis (HR 0.46, 95% CI 0.39-0.53, reference=BRACE) was strongly resistant to the influence of MRI lesion type and frequency with only marginal changes in the adjusted Hazard Ratio point estimate and associated confidence interval observed (adjusted Hazard Ratio: aHR 0.47, 95% CI 0.35-0.62 adjusted for ≥1 T1 Gd+ lesions; aHR 0.48, 95% CI 0.37, 0.62 adjusted for ≥9 T2 hyperintense lesions and aHR 0.46, 95% CI 0.31-0.68 adjusted for all of ≥1 T1 Gd+, ≥9 T2 hyperintense, ≥1 infratentorial, ≥1 juxtacortical and ≥2 periventricular lesions). A similar preservation of the effect size estimated in the primary analysis was further observed when both the time to switch therapy discontinuation and disability progression models were extended to adjust for baseline MRI data. The MRI-adjusted model of both the treatment discontinuation and disability progression models returned adjusted HR point estimates that differed by only 0.01 (treatment discontinuation: aHR 0.39. 95% CI 0.26-0.59 compared with a primary analysis estimate of HR 0.40, 95% CI 0.34-0.47; disability progression: aHR 0.75, 95% CI 0.54-0.98 adjusted model, HR 0.74, 95% CI 0.55-0.98 primary analysis).

Using post-hoc Rosenbaum sensitivity analyses of our propensity-matched relapse, persistence and progression models we estimated that an unobserved confounder would need to produce a minimum 3.01, 2.75 and 3.18 fold increase in the rate of relapse, discontinuation and progression respectively in order to reject the inference of a treatment effect in favour of selection effects. These represent improbably large differences in the context of the point estimates and associated...
confidence intervals observed, thus we can conclude that our observation of efficacy and persistence differentials by switch treatment arm are reasonably robust to unmeasured influences.

**DISCUSSION**

There are no head-to-head randomised clinical trials comparing efficacy outcomes between patients who switch to natalizumab or to another BRACE therapy after a BRACE treatment failure, so that this question can currently only be examined in real-world datasets, including cohort studies such as MSBase and TOP. Using propensity score matched samples from MSBase registry and TOP observational dataset, we demonstrated that superior outcomes were achieved in patients who switched to natalizumab after an on-treatment relapse. Compared to patients who switched between BRACE therapies, patients who switched to natalizumab had significantly reduced risks of further relapse occurrence, treatment discontinuation, and disability progression. Cumulative 24-month total disability burden, assessed using an exploratory AUC analysis, was also significantly lower in patients who switched to natalizumab. Subgroup analyses demonstrated that similar relapse and treatment discontinuation results are observed regardless of the type of prior BRACE treatment. Whilst the relative reduction in ARR associated with natalizumab was similar across all three prior treatment sub-groups (ranging from 65% to 75%), the comparative hazard of first relapse on switch therapy associated with natalizumab relative to BRACE was greatest in those patients previously exposed to both IFNβ and glatiramer acetate (Figure 4). This same sub-group also recorded the greatest differential in switch therapy discontinuation rates across all three sub-groups (Figure 5). This suggests there may be identifiable subsets of patients who response differently to IFNβ and/or glatiramer acetate. However, regardless of the identity or sequence of prior BRACE exposure, natalizumab treatment initiation was observed to be consistently associated with an efficacy advantage relative to BRACE switches.
Smaller studies have demonstrated results broadly consistent with these findings. Lanzillo and colleagues performed an adjusted analyses of 12-month outcomes after natalizumab or Rebif® initiation (n=42/group, mixed pre-treated and treatment-naïve at baseline) and showed better relapse, MRI, and disability outcomes in the natalizumab group. In a prospective, observational propensity-score adjusted but unmatched study of 267 relapsing-remitting MS patients in two Italian centers, Prosperini et al. demonstrated that patients who escalated to natalizumab (n=106) after first-line treatment failure on IFNβ or GA had a significantly higher probability of remaining relapse free, disability progression free, and MRI activity free over the next 24 months after switching treatment compared to patients who switched between IFNβ and/or GA formulations (n=161). These investigators controlled for some of the imbalance in baseline disease characteristics by deriving a propensity score of treatment assignation and adjusting for it, but a propensity matching approach such as the one detailed here has previously been demonstrated to provide superior control of confounding factors and thus superior attribution of treatment benefit compared to non-matched multivariable regression models. This is true even when regression models are fully adjusted for all available baseline characteristics. Unlike randomized controlled trials, propensity score based approaches cannot adjust for confounder imbalance in baseline characteristics that have not been recorded. However, propensity matching as a variety of pseudo-randomisation has been demonstrated to both reduce selection bias and closely approximate the risk estimates derived from randomized trials, including in the MSBase dataset.

The AUC serial disability time plots, albeit exploratory, were attempted to capture and estimate the total burden of complex disability trajectories commonly observed for relapsing MS patients in a clinical practice setting. Compared with the more commonly used summary measures such as the time to first confirmed disability progression analysis, which we have also presented herein, an AUC analysis arguably permits better estimation of a patient’s total study duration experience with respect to disability progression and thus superior attribution of any differences observed between

274
treatment groups. Since these serial disability plots explicitly attempt to capture and quantify the total changes in a patient’s on-treatment course, the use of an AUC disability metric is proposed to be more clinically meaningful than summary measures of EDSS change. In our analyses, treatment persistence was markedly longer after switch to natalizumab then switch to IFNβ/GA. These results are consistent with an analysis of US claims database which also showed that patients who switched to natalizumab demonstrated greater treatment persistence compared to those who switched to an alternative DMT.39

While this study focused on treatment efficacy and persistence, these are not the only important factors for clinicians to consider when weighing these treatment options. Treatment safety, in particular the risk of natalizumab-associated progressive multifocal leukoencephalopathy (PML), must also be assessed. Although not addressed here, evaluation of comparative natalizumab treatment benefits in clinical practice needs to be balanced with appropriate risk-stratification for PML, including testing for JC virus antibody status, to optimize informed and personalized treatment decisions.

The strongest results of our study concern the efficacy differentials observed in time to first relapse and treatment persistence favouring natalizumab. Although a comparable advantage was further observed with regards to confirmed disability progression, this result is less robust as it applies to that subset of the larger eligible sample who recorded a minimum of three prospective EDSS assessments. Furthermore our study only considers switches from BRACE to natalizumab and not alternate switch scenarios such as natalizumab to newer era oral DMDs. Generalisation of the efficacy advantages observed in this study for patients who switched to natalizumab may be limited by the characteristics of this patient population (most were recruited from large tertiary MS centres) and by the potential for treatment indication bias that may not have been adjusted for in the matched datasets. Propensity score matching was employed in this study to eliminate or reduce
known or suspected confounders of treatment allocation. However, unlike true randomisation, propensity matching cannot eliminate confounding secondary to imbalance of unknown or unmeasured confounders, and this remains a major limitation of this study. However, it would be expected that any residual bias would not favour the natalizumab cohort, as the known baseline variables in the unmatched populations indicated that those switching to natalizumab had much worse disease. Whilst subgroup and sensitivity analyses consistently demonstrated comparable efficacy advantages for patients who switched to natalizumab, future analyses adjusting for an expanded set of baseline characteristics, such as lesion number and distribution on baseline cerebral MRI, would be useful to corroborate these observations.

In the absence of randomized clinical trials, propensity-matching techniques can estimate the benefits associated with various treatment decisions in a clinical practice setting. Using a large real-world dataset, we have shown that patients who relapse on BRACE therapies have better outcomes if they switch to natalizumab rather than switching to another BRACE therapy. This extends to relapse rates, treatment persistence and, in the largest cohort examined, rates of first disability progression events.

**Funding**

The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Early Career Fellowship [1071124], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. The TYSABRI Observational Program (TOP) is fully funded by Biogen Idec. Biogen Idec provided writing assistance for the development of this manuscript. All statistical analyses were performed independently by the MSBase Foundation statistician, Tim Spelman.
ACKNOWLEDGEMENTS

MSBase study group co-investigators and contributors: From the MS-Centrum Nijmegen, Nijmegen, The Netherlands Dr Cees Zwanikken; from Hospital S. Joao, Porto, Portugal, Maria Edite Rio; Veszprem Megyei Csölnoky Ferenc Korhaz, Veszprem, Hungary, Dr Imre Piroska; from Jewish General Hospital, Montreal, Canada, Dr Fraser Moore; from Josa Andras Hospital, Nyiregyhaza, Dr Tunde Erdelyi; The Alfred Hospital and Monash University, Melbourne, Australia, Dr Olga Skibina; from Cliniques Universitaires Saint-Luc, Brussels, Belgium, Dr Vincent Van Pesch; from Ospedali Riuniti di Salerno, Salerno, Italy, Dr Gerardo Iuliano; from Jeroen Bosch Ziekenhuis, Den Bosch, The Netherlands, Dr Erik van Munster; from FLENI, Buenos Aires, Argentina, Dr Marcela Fiol, Dr Jorge Correale and Dr Celica Ysraaelit; from Department NEUROFARBA, Section of Neurosciences, University of Florence, Florence, Italy, Dr Maria Pia Amato; from Francicus Ziekenhuis, Roosendaal, The Netherlands, Dr Leontien den Braber-Moerland; from New York University Langone Medical Center, New York, USA, Dr Joseph Herbert and Dr Iliya Kister; from Hopital Tenon, Paris, France, Dr Etienne Roullet; from Jahn Ferenc Teaching Hospital, Budapest, Hungary, Dr Krisztian Kasa; from Central Clinical Emergency Military Hospital, Bucharest, Romania, Dr Carmen-Adella Sirbu; from the Geelong Hospital, Geelong, Australia, Dr Cameron Shaw; from HIGA Gral. San Martin, La Plata, Argentina, Dr Santiago Vetere; from the Westmead Hospital, Sydney, Australia, Dr Steve Vucic; from the Clinic of Neurology Clinical Center, Skopje, Macedonia, Dr Tatjana Petkovska-Boskova; from the Bombay Hospital Institute of Medical Sciences, Mumbai, India, Dr Bhim Singhal; from the Instituto de Neurociencias, Cordoba, Argentina, Dr Elizabeth Alejandra Bacile Bacile; from the Hospital Ecoville, Brazil, Dr Walter Oleschko Arruda; from the Centre hospitalier del’Universite de Montreal, Hopital Notre-Dame, Canada, Dr Elaine Roger and Dr Pierre Despault; from the Royal Melbourne Hospital, Australia, Dr Mark Marriott, Dr Anneke Van der Walt, Dr John King, Dr Jill Byron and Ms Lisa Morgan; from Box Hill Hospital, Monash University, Australia, Ms Jodi Haartsen; from Department of Neuroscience and Imaging, University ‘G. d’Annunzio’, Italy, Dr Giovanna De Luca, Dr
Timothy Denis Spelman (58172)

Valeria Di Tommaso, Dr Daniela Travaglini, Dr Erika Pietrolongo, Dr Maria di Ioia, Dr Deborah Farina and Dr Luca Mancinelli; from Hospital Italiano, Argentina, Dr Juan Ignacio Rojas and Dr Liliana Patrucco; from Ospedale di Macerata, Italy, Dr Elisabetta Cartechini and Dr Giorgio Giuliani; from John Hunter Hospital, Australia, Dr David Williams and Dr Lisa Dark; from Buenos Aires, Argentina, Dr Aldo Savino; and from Sheba Medical Center, Tel Hashomer, Israel, Dr Joab Chapman; from Assaf Harofeh Medical Center, Beer-Yaakov, Israel, Dr Shlomo Flechter; from Hospital Italiano, Buenos Aires, Argentina, Dr Edgardo Cristiano; from Centro Internacional de Restauracion Neurologica, Havana, Cuba. Dr Jose Antonio Cabrera-Gomez; from INEBA, Buenos Aires, Argentina, Dr Maria Laura Saladino; from Hospital Fernandez, Buenos Aires, Argentina, Dr Norma Deri; from Craigavon Area Hospital, Portadown, UK, Dr Orla Gray; from St Vincent’s Hospital, Melbourne, Australia; Dr Mark Paine; and from Mater Dei Hospital, Malta; Dr Norbert Vella; Mr Samir Méchati, Mr Eric Bianchi, Mr Alexandru Bulla and Mr Matthieu Corageoud. No compensation has been received for the persons who have made substantial contributions to the work but do not qualify as authors. Nolan Campbell of Biogen Idec provided editorial assistance which involved reference management, coordination of author feedback and manuscript versioning management. Patrick Campbell of Biogen Idec provided minor graphic design assistance with a subset of figures during the development of this manuscript.

Author contributions:

T Spelman conceptualised and designed the study, conducted and interpreted the analysis and drafted, revised and approved the manuscript. H Butzkueven conceptualised the study and drafted, revised and approved the manuscript. T Kalincik and S Belachew conceptualised the study and revised and approved the manuscript. A Zhang, F. Pellegrini, H Wiendl, L Kappos, L Tsvetkova, R Hyde, F Verheul, F Grand-Maison, G Izquierdo, P Grammond, P Duquette, A Lugaresi, J Lechner-Scott, C Oreja-Guevara, R Hupperts, T Petersen, M Barnett and M Trojano contributed substantially to data acquisition, interpretation of the analysis and have revised and approved the manuscript.
Disclosures:

**Alessandra Lugaresi** is a Bayer Schering, Biogen Idec, Genzyme/Sanofi, Merck Serono Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi and Teva and research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva. Prof Lugaresi has also received travel and research grants from the Associazione Italiana Sclerosi Multipla and was a Consultant of “Fondazione Cesare Serono”.

**Annie Zhang** is an employee of Biogen Idec Inc.

**Celia Oreja-Guevara** reports no conflicts of interest.

**Fabio Pelligrini** received honoraria for speaking and personal compensation for consulting services from Biogen Idec.

**Francois Grand’Maison** received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.

**Freek Verheul** reports no conflicts of interest.

**Guillermo Izquierdo** received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

**Heinz Wiendl** received compensation for serving on scientific advisory boards for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi; speaker honoraria and travel support from Bayer Schering AG, Bayer Vital GmbH, Biogen Idec, CSL Behring, Fresenius Medical Care, Genzyme, GlaxoSmithKline, GW, Merck Serono, Novartis, and Sanofi; compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi; research support from Bayer Vital GmbH, Biogen Idec, Merck Serono, Novartis, Sanofi Germany, and Sanofi US.

**Helmut Butzkueven** received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck
Timothy Denis Spelman (58172)

Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.

Jeannette Lechner Scott reports no conflicts of interest.

Larisa Tsvetkova is an employee of Biogen Idec Inc.

Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

Michael Barnett has received honoraria for participation in advisory boards and travel sponsorship from Novartis, BioCSL, Genzyme and Biogen Idec.

Pierre Duquette has received honoraria for organising CME events and has obtained funding to attend meetings from Biogen Idec, EMD Serono, TEVA Neuroscience, Novartis, and Genzyme, has received funding for investigator-initiated trials with Biogen Idec, EMD Serono and Novartis, and has received peer-review funding from CIHR and from the MS Society of Canada.

Pierre Grammond is a Novartis, Teva-neuroscience, Biogen Idec advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis.

Raymond Hupperts received honoraria as consultant on scientific advisory boards from Merck-Serono, Biogen-Idec, Genzyme-Sanofi and Teva, research funding from Merck-Serono and Biogen-Idec, and speaker honoraria from Sanofi-Genzyme.

Robert Hyde is an employee of Biogen Idec Inc.

Shibeshih Belachew is an employee of Biogen Idec Inc.

Thor Petersen received funding or speaker honoraria from Biogen Idec, Merck Serono, Novartis, Bayer Schering, Sanofi-Aventis, Roche, and Genzyme.

Tim Spelman received honoraria for consultancy and funding for travel from Biogen Idec Inc.
Tomas Kalincik received compensation for travel from Novartis, Biogen Idec, Sanofi Aventis, Teva and Merck Serono.

REFERENCES


Supplemental Figure 1: Distribution of propensity scores by treatment arm prior to (A) and after (B) propensity matching.
## Supplementary Table 1: Baseline characteristics of unmatched patients by switch group

<table>
<thead>
<tr>
<th>Variable (at baseline)</th>
<th>Total excluded population post-matching</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NTZ</td>
<td>IFNβ/GA</td>
</tr>
<tr>
<td>n</td>
<td>2508</td>
<td>278</td>
</tr>
<tr>
<td>Female, %</td>
<td>69.5</td>
<td>85.4</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>37 (29, 44)</td>
<td>35 (29, 43)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.497</td>
<td></td>
</tr>
<tr>
<td>Disease duration, median (IQR)</td>
<td>6.9 (3.5, 11.8)</td>
<td>5.1 (2.3, 9.5)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>Proportion disease duration on DMT, median (IQR)</td>
<td>0.6 (0.4, 0.8)</td>
<td>0.2 (0.1, 0.3)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Number of DMT starts mean (SD)</td>
<td>1.6 (0.8)</td>
<td>1.2 (0.6)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Number of DMT starts / disease duration, mean (SD)</td>
<td>0.3 (0.4)</td>
<td>0.4 (0.5)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.438</td>
<td></td>
</tr>
<tr>
<td>Baseline EDSS, median (IQR)</td>
<td>3.5 (2, 4.5)</td>
<td>2 (1, 2.5)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Total relapse onsets last 12 months, mean (SD)</td>
<td>2.1 (1.1)</td>
<td>1.5 (0.7)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Total steroid-treated relapses last 12 months, mean (SD)</td>
<td>1.9 (1.1)</td>
<td>0.3 (0.6)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Total relapse onsets last 24 months, mean (SD)</td>
<td>3.1 (1.6)</td>
<td>2.2 (1.3)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Total steroid-treated relapses last 24 months, mean (SD)</td>
<td>2.7 (1.5)</td>
<td>0.3 (0.6)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
### Title
Comparative efficacy of first-line natalizumab versus IFNβ or glatiramer acetate in relapsing MS (Accepted paper – Neurology: Clinical Practice)

<table>
<thead>
<tr>
<th>Title</th>
<th>Comparative efficacy of first-line natalizumab versus IFNβ or glatiramer acetate in relapsing MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Tim Spelman¹⁺ (MBBS), Tomas Kalincik¹ (PhD), Vilija Jokubaitis¹ (PhD), Annie Zhang² (PhD), Fabio Pellegrini² (PhD), Heinz Wiendl⁴ (MD), Shibishih Belachew² (PhD, MD), Robert Hyde² (PhD), Freek Verheul⁵ (MD), Alessandra Lugaresi⁶ (MD), Eva Havrdova⁷ (PhD), Dana Horakova⁷ (MD), Pierre Grammond⁸ (MD), Pierre Duquette⁹ (MD), Alexandre Prat⁹ (MD), Gerardo Iuliano¹⁰ (MD), Murat Terzi¹¹ (MD), Guillermo Izquierdo¹² (MD), Raymond Hupperts¹³(PhD), Cavit Boz¹⁴ (MD), Eugenio Pucci¹⁵ (MD), Giorgio Giuliani¹⁵ (MD), Patrizia Sola¹⁶ (MD), Daniele La Spitaleri¹⁷ (MD), Jeannette Lechner-Scott¹⁸ (MD), Roberto Bergamaschi¹⁹ (MD), Francois Grand-Maison²⁰ (MD), Franco Granella²¹ (MD), Ludwig Kappos²² (MD), Maria Trojano<em>²³(MD), Helmut Butzkueven</em>¹,²² (PhD) on behalf of the MSBase Investigators and the TOP investigators.</td>
</tr>
<tr>
<td>Author information</td>
<td>¹ Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia</td>
</tr>
<tr>
<td></td>
<td>² Biogen Idec Inc., Cambridge, MA</td>
</tr>
<tr>
<td></td>
<td>³ Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy</td>
</tr>
<tr>
<td></td>
<td>⁴ Department of Neurology, University of Münster, Münster, Germany</td>
</tr>
<tr>
<td></td>
<td>⁵ Groene Hart Ziekenhuis, Gouda, The Netherlands</td>
</tr>
<tr>
<td></td>
<td>⁶ MS Center, Department of Neuroscience, Imaging and Clinical Sciences, University ‘G. d’Annunzio’, Chieti, Italy</td>
</tr>
<tr>
<td></td>
<td>⁷ MS Center, Department of Neurology, First Medical Faculty, Charles University, Prague, Czech Republic</td>
</tr>
<tr>
<td></td>
<td>⁸ Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada</td>
</tr>
<tr>
<td></td>
<td>⁹ Hôpital Notre Dame, Montreal, Canada</td>
</tr>
<tr>
<td></td>
<td>¹⁰ Ospedali Riuniti di Salerno, Salerno, Italy</td>
</tr>
<tr>
<td></td>
<td>¹¹ 19 Mayis University, Medical Faculty, Turkey</td>
</tr>
<tr>
<td></td>
<td>¹² Hospital Universitario Virgen Macarena, Sevilla, Spain</td>
</tr>
<tr>
<td></td>
<td>¹³ Orbis Medical Centre, Sittard-Geleen, The Netherlands</td>
</tr>
<tr>
<td></td>
<td>¹⁴ KTU Medical Faculty Farabi Hospital, Trabzon, Turkey</td>
</tr>
<tr>
<td></td>
<td>¹⁵ Neurology Unit, ASUR Marche – AV3, Macerata, Italy.</td>
</tr>
<tr>
<td></td>
<td>¹⁶ Nuovo Ospedale Civile S.Agostino, Modena, Italy</td>
</tr>
<tr>
<td></td>
<td>¹⁷ AORN San Giuseppe Moscati, Avellino, Italy</td>
</tr>
<tr>
<td></td>
<td>¹⁸ John Hunter Hospital, Newcastle, Australia</td>
</tr>
<tr>
<td></td>
<td>¹⁹ Neurological Institute IRCCS Mondino, Pavia, Italy</td>
</tr>
<tr>
<td></td>
<td>²⁰ University of Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada</td>
</tr>
<tr>
<td></td>
<td>²¹ University of Parma, Parma, Italy</td>
</tr>
<tr>
<td></td>
<td>²² Department of Neurology, University Hospital Basel, Basel, Switzerland</td>
</tr>
<tr>
<td></td>
<td>²³ Dept of Neurology, Eastern Health, Monash University, Box Hill ,</td>
</tr>
</tbody>
</table>
| **Contact information** | Dr Tim Spelman  
Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia  
tim@burnet.edu.au  
Ph: 61 3 9342 4406  
Fx: 61 3 9349 5997 |
|-------------------------|--------------------------------------------------|
| **Word Count**          | Abstract: 249  
Text: 3437 |
| **Character Count (with spaces)** | Title: 96 |
| **Study funding**       | The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Early Career Fellowship [1071124], NHMRC Project Grants [1032484 and 1083539], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL. The TYSABRI Observational Program (TOP) is fully funded by Biogen. E Havrdova and D Horakova have been supported by Research Grant of Czech Ministry of |
Author contributions:

T Spelman conceptualized and designed the study, conducted the statistical analysis, interpreted the analysis and drafted and revised the paper

T Kalincik conceptualized and designed the study, interpreted the analysis and revised the paper

V Jokubaitis interpreted the analysis and revised the paper

A Zhang conceptualized and designed the study and revised the paper

F Pelligrini conceptualized and designed the study and revised the paper

H Wiendl interpreted the analysis and revised the paper

S Belachew conceptualized and designed the study, interpreted the analysis and revised the paper

R Hyde conceptualized and designed the study, interpreted the analysis and revised the paper

F Verheul interpreted the analysis and revised the paper

A Lugaresi interpreted the analysis and revised the paper

E Havrdova interpreted the analysis and revised the paper

D Horakova interpreted the analysis and revised the paper

P Grammond interpreted the analysis and revised the paper

P Duquette interpreted the analysis and revised the paper

A Prat interpreted the analysis and revised the paper

G Iuliano interpreted the analysis and revised the paper

M Terzi interpreted the analysis and revised the paper

G Izquierdo interpreted the analysis and revised the paper

R Hupperts interpreted the analysis and revised the paper

C Boz interpreted the analysis and revised the paper

E Pucci interpreted the analysis and revised the paper

G Giuliani interpreted the analysis and revised the paper

P Sola interpreted the analysis and revised the paper

D La Spitaleri interpreted the analysis and revised the paper
J Lechner-Scott interpreted the analysis and revised the paper
R Bergamaschi interpreted the analysis and revised the paper
F Grand-Maison interpreted the analysis and revised the paper
F Granella interpreted the analysis and revised the paper
L Kappos interpreted the analysis and revised the paper
M Trojano interpreted the analysis and revised the paper
H Butzkueven conceptualized and designed the study, interpreted the analysis and revised the paper

Author disclosures:

T Spelman received honoraria for consultancy and funding for travel from Biogen Idec Inc.
T Kalincik received compensation for conference travel and consultancy/speaker honoraria from Novartis, Biogen Idec, Genzyme, Sanofi Aventis, Teva, BioCSL and Merck Serono
V Jokubaitis has received conference travel support from Novartis.
A Zhang is an employee of Biogen Idec Inc.
F Pelligrini is an employee of Biogen Idec Inc.
H Wiendl received compensation for serving on scientific advisory boards for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi; speaker honoraria and travel support from Bayer Schering AG, Bayer Vital GmbH, Biogen Idec, CSL Behring, Fresenius Medical Care, Genzyme, GlaxoSmithKline, GW, Merck Serono, Novartis, and Sanofi; compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi; research support from Bayer Vital GmbH, Biogen Idec, Merck Serono, Novartis, Sanofi Germany, and Sanofi US.
S Belachew is an employee of Biogen Idec Inc.
R Hyde is an employee of Biogen Idec Inc.
F Verheul is an advisory board member for Teva Biogen Merck Serono and Novartis.
A Lugaresi was a Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Merck Serono,
Timothy Denis Spelman (58172)

Novartis, Sanofi and Teva, research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, travel and research grants from the Associazione Italiana Sclerosi Multipla.

**E Havrdova** received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

**D Horakova** received speaker honoraria and consulting fees from Biogen Idec, Merck Serono, Teva and Novartis, as well as support for research activities from Biogen Idec.

**P Grammond** is a Novartis, Teva-neuroscience, Biogen Idec advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis.

**P Duquette** has received honoraria for organising CME events and has obtained funding to attend meetings from Biogen Idec, EMD Serono, TEVA Neuroscience, Novartis, and Genzyme, has received funding for investigator-initiated trials with Biogen Idec, EMD Serono and Novartis, and has received peer-review funding from CIHR and from the MS Society of Canada.

**A Prat** reports no conflicts of interest.

**G Iuliano** had travel/accommodations/meeting expenses funded by Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi Aventis, and Teva

**M Terzi** received travel grants from Merck Serono, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.

**G Izquierdo** received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

**R Hupperts** received honoraria as consultant on scientific advisory boards from Merck-Serono, Biogen-Idec, Genzyme-Sanofi and Teva, research funding from Merck-Serono and Biogen-Idec, and speaker honoraria from Sanofi-Genzyme.

**C Boz** received conference travel support from Biogen Idec, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.
Timothy Denis Spelman (58172)

E Pucci served on scientific advisory boards for Genzyme, Novartis and Biogen-Idec; he has received honoraria and travel grants from Sanofi Aventis, Novartis, Biogen Idec, Merck Serono, Genzyme and Teva; he has received travel grants from Associazione Marchigiana Sclerosi Multipla e altre malattie neurologiche.

G Giuliani did not declare any competing interests.

P Sola did not declare any competing interests.

D La Spitaleri received honoraria as a consultant on scientific advisory boards by Bayer-Schering, Novartis and Sanofi-Aventis and compensation for travel from Novartis, Biogen Idec, Sanofi Aventis, Teva and Merck-Serono.

J Lechner-Scott has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen Idec, CSL, Genzyme Sanofi, Merck Serono and Novartis.

R Bergamaschi received speaker honoraria from Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi-Aventis, Teva; research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi-Aventis, Teva; congress and travel/accommodation expense compensations by Almirall, Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi-Aventis, Teva.

F Grand-Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.

F Granella did not declare any competing interests.

L Kappos received research support from Acorda, Actelion, Allozyne, BaroFold, Bayer HealthCare, Bayer Schering, Bayhill Therapeutics, Biogen Idec, Elan, European Union, Genmab, Gianni Rubatto Foundation, GlaxoSmithKline, Glenmark, MediciNova, Merck Serono, Novartis, Novartis Research Foundation, Roche, Roche Research Foundation, Sanofi-Aventis, Santhera, Swiss MS Society, Swiss National Research Foundation, Teva Neuroscience, UCB, and Wyeth.
Timothy Denis Spelman (58172)

**M Trojano** received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

**H Butzkueven** received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.
ABSTRACT

Objective: To compare efficacy and treatment persistence in treatment naïve patients with relapsing-remitting multiple sclerosis (RRMS) initiating natalizumab compared with interferon-beta (IFNβ)/glatiramer acetate (GA) therapies, using propensity score-matched cohorts from observational MS registries.

Methods: The study population initiated IFNβ/GA in the MSBase Registry or natalizumab in the TYSABRI Observational Program (TOP), had ≥3 months of on-treatment follow-up, and active RRMS, defined as ≥1 gadolinium-enhancing lesion on cerebral MRI at baseline and/or ≥1 relapse within the 12 months prior to baseline. Baseline demographics and disease characteristics were balanced between propensity-matched groups. Annualized relapse rate (ARR), time to first relapse, treatment persistence, and disability outcomes were compared between matched treatment arms in the total population (n=366/group) and subgroups with higher baseline disease activity.

Results: First-line natalizumab was associated with a 68% relative reduction in ARR from a mean (SD) of 0.63 (0.92) on IFNβ/GA to 0.20 (0.63) (p(signed-rank)<0.0001), a 64% reduction in the rate of first relapse (hazard ratio [HR]= 0.36; 95% confidence interval [CI]=0.28-0.47; p<0.001), and a 27% reduction in the rate of discontinuation (HR=0.73, 95%CI=0.58-0.93; p=0.01), compared with first-line IFNβ/GA therapy. Confirmed disability progression and area under the EDSS-time curve analyses were not statistically significant. Similar relapse and treatment persistence results were observed in each of the higher disease activity subgroups.

Conclusions: This study provides class IV evidence that first-line natalizumab for RRMS improves relapse and treatment persistence outcomes, compared to first-line IFNβ/GA. This needs to be balanced against the risk of Progressive Multifocal Leukoencephalopathy (PML) in natalizumab treated patients.
Classification of Evidence: This study provides class IV evidence that first-line natalizumab for RRMS improves relapse rates and treatment persistence outcomes, compared to first-line IFN/GA.
INTRODUCTION

In the pivotal, phase II AFFIRM trial, natalizumab demonstrated high efficacy in patients who were mostly (90%) naïve to disease modifying therapy (DMT).\(^1\) Despite this, in clinical practice natalizumab is often recommended for MS patients with inadequate response to other treatments or patients with high levels of disease activity.\(^2\) For much of the world the indicated use of natalizumab as a first-line therapy is restricted to patients with ≥2 relapses within one year and ≥1 gadolinium (Gd⁺)-enhancing lesion or a significant increase in T2 lesions on MRI.\(^3\) In settings where first-line use of natalizumab is not restricted, it is important for physicians to balance the potential efficacy benefits of natalizumab against a patient’s risk of progressive multifocal leukoencephalopathy (PML) before initiating natalizumab.\(^4\)

While factors that stratify natalizumab-associated PML risk, particularly anti-JC Virus antibody status, have been identified,\(^5\) the potential efficacy advantage of natalizumab specifically as first-line therapy over other DMTs has not been fully explored. Placebo-controlled trials such as AFFIRM\(^1\) and those comparing natalizumab as an adjunct therapy to IFNβ and GA versus IFNβ or GA monotherapy\(^6\) do not provide information on outcomes associated with initiating natalizumab monotherapy versus other treatment options commonly considered in clinical practice. There are no head-to-head clinical trials comparing the efficacy of first-line natalizumab treatment to other first-line DMTs.

In the respective pivotal clinical trials comparing active treatment to placebo, natalizumab monotherapy reduced ARR by 68%\(^5\)–\(^16\) compared to an around 30% reduction for interferon-beta preparations and glatiramer acetate. Although these trials largely enrolled treatment-naïve patients, they were conducted in different epochs resulting in different absolute annualised relapse rates (ranging from 0.73 to 1.28 in the placebo groups). A head-to-head comparison, is therefore needed to establish the comparative effectiveness of first-line natalizumab versus IFNβ or GA.
The primary objective of this study were to compare time to first relapse and treatment discontinuation in DMT naïve patients with active MS disease who initiated first-line natalizumab treatment compared with first-line Betaferon®/Betaseron®, Rebif®, Avonex®, Copaxone® or Extavia® (BRACE) treatments. Confirmed disability progression was studied as a secondary end-point. Propensity score-matching was employed to reduce the confounding inherent to observational studies. The study group has recently successfully employed this technique for MS treatment comparisons in the MSBase registry dataset and also in a combined MSBase/ TYSABRI Observational Program (TOP) dual dataset, two contemporaneous, real-world cohorts.

MATERIALS & METHODS

Standard Protocol Approvals, Registrations, and Patient Consents Patients treated with BRACE or natalizumab were extracted from the MSBase and Tysabri Observational Programme registries respectively. Details of these registries have been previously published. In both registries, relapse is characterised according to McDonald’s criteria, whilst interactive Neurostatus training is used for consistent assessment of EDSS. Project approval from an ethical standards committee on human experimentation (institutional or regional) for any experiments using human subjects was obtained by each clinical centre contributing data to either MSBase or TOP registries. Written informed consent was obtained from all patients contributing data to either registry in accordance to the local regulations and laws applicable at each clinic.

Study design

Patients and Subgroups

DMT naïve patients exhibiting disease activity (defined by ≥1 relapse within 12 months of baseline OR ≥1 Gd⁺-enhancing lesion at baseline) were eligible for analysis. Baseline was defined as the time of DMT commencement. Patients were censored either at the date of recording a first relapse,
discontinuation or progression event respectively for each analysis, otherwise the date of last clinician assessment point. Comparisons were made between propensity score-matched patients who initiated natalizumab versus BRACE therapy as their first-line DMT. Participants with <3 months of follow-up or no recorded cerebral MRI within 6 months prior to DMT commencement were excluded (Figure 1).

Subgroup analyses compared treatment efficacy and persistence in patients with various levels of baseline MS disease activity. The 3 subgroups considered were: 1) ≥1 relapse AND ≥1 Gd\(^+\)-enhancing lesion; 2) ≥2 relapses AND ≥1+ Gd\(^+\)-enhancing lesion; or 3) ≥1 relapse AND ≥1 Gd\(^+\)-enhancing lesion AND ≥9 T2 hyper-intense lesions. Matching treatment arms by propensity score was performed separately for the total eligible sample of the primary analysis and then repeated for each analysis subgroup (Figure 1).

**Figure 1: Study Profile**

Of the 5404 TOP participants at time of analysis, 518 were treatment naïve at baseline all of which recorded a minimum 24 months of pre-baseline follow-up whilst 430 of these also recorded at least 1 relapse in the 12 months prior to baseline or at least 1 Gd\(^+\) lesion on baseline MRI. Of the 30,417 enrolled MSBase participants, 11,564 were treatment naïve at baseline, 9018 recorded a minimum of 24 months pre-baseline activity and

---

<table>
<thead>
<tr>
<th>Registry participants at time of analysis</th>
<th>TOP</th>
<th>MSBase</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMT naïve, ≥24 months of disease history, ≥1 relapse in the prior 12 months OR ≥1 Gd(^+) enhancing lesion, initiating natalizumab or BRACE therapy*</td>
<td>5,404</td>
<td>30,417</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Higher disease activity subgroups</th>
<th>TOP</th>
<th>MSBase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis</td>
<td>430</td>
<td>2,120</td>
</tr>
<tr>
<td>≥1 relapse AND ≥1 Gd(^+) lesion</td>
<td>259</td>
<td>121</td>
</tr>
<tr>
<td>≥2 relapses AND ≥1 Gd(^+) lesion</td>
<td>198</td>
<td>75</td>
</tr>
<tr>
<td>≥1 relapse, AND ≥1 Gd(^+) lesion AND ≥9 T2 lesions</td>
<td>202</td>
<td>123</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Successfully matched by propensity score*</th>
<th>TOP</th>
<th>MSBase</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥3 EDSS score reports including baseline</td>
<td>366</td>
<td>366</td>
</tr>
<tr>
<td>On-therapy follow-up ≥34 months with EDSS reported within 16 months of month 24</td>
<td>193</td>
<td>193</td>
</tr>
<tr>
<td>Populations for analyses of ARR, time to first relapse, and time to treatment discontinuation</td>
<td>36</td>
<td>55</td>
</tr>
<tr>
<td>Population for analysis of time to 3-month confirmed EDSS progression</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Population for analysis of AUC</td>
<td>59</td>
<td>69</td>
</tr>
</tbody>
</table>

*Patients in TOP initiated natalizumab. Patients in MSBase initiated a BRACE therapy. *Propensity score matching was performed independently for the primary analysis and in each higher disease activity subgroup.
2120 of these also recorded at least 1 relapse in the 12 months prior to baseline or at least 1 Gd+ lesion on baseline MRI.

**Efficacy Measures**

The primary research question was whether there was a difference in the rate of first relapse or treatment discontinuation in patients on first line natalizumab relative to first-line BRACE. A classification of class IV evidence applies to both outcomes. Annualised relapse rate (ARR), time to first on-treatment relapse, and treatment persistence were analysed as primary outcomes. Secondary endpoints included time to 3-month confirmed disability progression and 24-month change in area under the serial disability/ time curve (AUC). Confirmed disability progression events were defined as a minimum 3-month confirmed increases of ≥0.5, ≥1.0 and ≥1.5 points for patients with baseline EDSS >5.5, between 1.0 and 5.5 and those with baseline EDSS of 0 respectively. EDSS scores recorded within 30 days post-relapse were excluded. The comparison of disability progression by treatment arm was limited to patients contributing at least 3 EDSS assessment points, as the minimum required to first observed and then confirm a progression event.

Serial disability/time AUC change comparisons were performed to estimate patients’ cumulative on treatment disability experience. EDSS scores assessed within a 24-month interval from baseline were initially plotted, and AUC calculated as previously described in our published reports. As a sensitivity analysis AUC change in EDSS was also calculated using the trapezoidal rule.

**Statistical analyses**

Data from the MSBase and TOP registries were aggregated according to a pre-specified protocol. Categorical variables were summarized using frequency and percentage and continuous factors
Timothy Denis Spelman (58172)

summarised using median and inter-quartile range (IQR) or mean and standard deviation (SD) as appropriate.

For both the primary analysis group and each of the three high disease activity sub-groups, patients from the natalizumab group were matched to a comparable patient in the BRACE arm using propensity matching. Sex, age, EDSS, disease duration, and the count of both total relapses and steroid-treated relapses in the 12 and 24 months prior to baseline were used to 1:1 match patients employing a 5:1 digit “greedy” matching algorithm as previously described by this study group. As a sensitivity analysis, clinic country was included for the derivation of the propensity score. A 1:1 match was preferred over a 1-to-many match as the latter introduced unacceptable imbalance secondary the poorer quality of the secondary and tertiary matches. Balance of baseline covariates by treatment arm post-matching was assessed via analysis of standardized differences and matched Wilcoxon signed-rank and McNemar tests. Wilcoxon rank-sum for continuous factors and a chi-square test for categorical variables were used to compare unmatched baseline characteristics by treatment arm.

A Cox Marginal Model was used to test for differences in time to first on-treatment relapse event, treatment persistence and disability progression by treatment arm. Scaled Schoenfeld residuals were used to test each model for underlying hazard proportionality. The potential influence of informative censoring secondary to group differences in follow up time on event ascertainment was studied by extending the Cox models to include follow-up differentials as competing risks for observing end-point events. A test for interaction was used to assess the sub-groups for treatment effects.

Quantile median regression was used to compare AUC change in EDSS across treatment arms adjusting for the matched pairs via Censored Least Absolute Deviations (CLAD).
extension of the Cochran-Armitage test\textsuperscript{28} was used to check for non-linearity in associations between AUC change in EDSS and treatment arm. An a priori specified Rosenbaum sensitivity analyses was applied post-estimation across all outcomes to test the sensitivity of the matched models to unobserved heterogeneity secondary to baseline characteristics that were either not collected or incompletely observed.\textsuperscript{29} As a sensitivity analysis, the relapse, discontinuation and progression models were re-run using first-line natalizumab initiations sourced from within MSBase. In all analyses, \( p < 0.05 \) was considered significant. All analyses were conducted in Stata version 13 (StataCorp, College Station, Texas).

**RESULTS**

**Patients**

Of the eligible patients (Figure 1, natalizumab, 430; BRACE, 2120), 366 (85.1\%) first-line natalizumab commencements were successfully matched to a first-line BRACE initiation. Significant imbalance in pre-matching baseline covariates were observed across treatment arms; patients commencing natalizumab were younger, had shorter disease duration, higher median EDSS, and greater pre-treatment relapse activity compared to those commencing BRACE (Table 1). Conversely, there was no significant observed imbalance between treatment arms after matching (Table 2). Mean (SD) relapses in the 12 months prior to treatment initiation was 1.9 in both arms whilst median baseline was 3 across both arms, representing a relatively active and severe disease cohort. Mean (SD) on-treatment follow-up was 3.1 years (2.7) in the BRACE group compared to 2.0 years (1.4) in the first-line natalizumab treatment arm (\( p(\text{signed-rank}) = 0.001 \)). Mean (SD) time between on-treatment assessments was 5.9 months (5.2) on first-line BRACE therapy relative to the 6.4 months (2.7) observed in the natalizumab arm (\( p(\text{signed-rank}) = 0.103 \)). Similarly there was no difference in the median on-study visit density between the matched groups with the natalizumab group recording a
Table 1: Baseline characteristics of unmatched patients

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Natalizumab (n=430)</th>
<th>BRACE (n=2120)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean (IQR)</td>
<td>33 (27, 42)</td>
<td>37 (30, 44)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sex, % female</td>
<td>70</td>
<td>72</td>
<td>0.351</td>
</tr>
<tr>
<td>Disease duration, years, mean (SD)</td>
<td>4.2 (4.0)</td>
<td>7.9 (6.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>EDSS, median (IQR)</td>
<td>3.0 (2.0, 4.0)</td>
<td>2.0 (1.5, 3.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Relapse onsets in prior year, mean (SD)</td>
<td>2.1 (1.0)</td>
<td>1.4 (0.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Relapse onsets in prior 2 years, mean (SD)</td>
<td>2.5 (1.2)</td>
<td>1.8 (1.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Steroid-treated relapses in prior year, mean (SD)</td>
<td>1.5 (1.0)</td>
<td>0.6 (0.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Steroid-treated relapses in prior 2 years, mean (SD)</td>
<td>1.7 (1.2)</td>
<td>0.7 (0.8)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviations: EDSS = Expanded Disability Status Scale; IQR = interquartile range; SD = standard deviation; BRACE = Betaferon®, Betaseron®, Rebif®, Avonex®, Copaxone® or Extavia®

Table 2: Baseline characteristics of propensity score-matched patients

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Natalizumab (n=366)</th>
<th>BRACE (n=366)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean (IQR)</td>
<td>34 (23, 42)</td>
<td>35 (29, 43)</td>
<td>0.415</td>
</tr>
<tr>
<td>Sex, % female</td>
<td>69</td>
<td>70</td>
<td>0.673</td>
</tr>
<tr>
<td>Disease duration, years, mean (SD)</td>
<td>4.7 (4.3)</td>
<td>5.6 (4.1)</td>
<td>0.329</td>
</tr>
<tr>
<td>EDSS, median (IQR)</td>
<td>3.0 (2.0, 4.0)</td>
<td>3.0 (2.0, 4.0)</td>
<td>0.338</td>
</tr>
<tr>
<td>Relapse onsets in prior year, mean (SD)</td>
<td>1.9 (0.8)</td>
<td>1.9 (1.0)</td>
<td>0.272</td>
</tr>
<tr>
<td>Relapse onsets in prior 2 years, mean (SD)</td>
<td>2.3 (1.0)</td>
<td>2.3 (1.2)</td>
<td>0.591</td>
</tr>
<tr>
<td>Steroid-treated relapses in prior year, mean (SD)</td>
<td>1.3 (0.8)</td>
<td>1.3 (0.9)</td>
<td>0.810</td>
</tr>
<tr>
<td>Steroid-treated relapses in prior 2 years, mean (SD)</td>
<td>1.4 (0.9)</td>
<td>1.5 (1.1)</td>
<td>0.699</td>
</tr>
</tbody>
</table>

Abbreviations: EDSS = Expanded Disability Status Scale; IQR = interquartile range; SD = standard deviation; BRACE = Betaferon®, Betaseron®, Rebif®, Avonex®, Copaxone® or Extavia®

*All post-matching baseline characteristics had a standardized difference between -10% and 10%.
median (IQR) number of visits per year of 2.41 (1.94, 3.09) compared with 2.57 (1.91, 3.85) in the BRACE arm \( (p_{\text{signed-rank}}=0.142) \).

**On-treatment relapse**

On-treatment ARR was significantly lower (relative reduction, 68%; \( p_{\text{signed-rank}}<0.0001 \)) for patients who initiated natalizumab (mean, 0.20; SD, 0.63) as their first-line DMT compared to BRACE (mean, 0.63; SD, 0.92) (Table 3). During the study, 76 (20.7%) natalizumab patients experienced an on-treatment relapse at a rate of 12.7 first relapse events per 100 person-years; 224 (61.2%) BRACE patients experienced an on-treatment relapse at a rate of 25.1 first relapse events per 100 person-years of on-treatment follow-up, corresponding to a 64% reduction in the relapse rate for patients who initiated natalizumab (hazard ratio [HR]=0.36, 95% confidence interval [CI]=0.28-0.47) (Figure 2A).

![Figure 2: Time to A) first relapse and (B) treatment discontinuation on first-line treatment](image)
Table 3: Summary of propensity-matched treatment group outcomes by prior disease activity

<table>
<thead>
<tr>
<th></th>
<th>≥1 Relapse OR</th>
<th>≥1 Relapse AND ≥1 Gd⁺ lesion</th>
<th>≥2 Relapses AND ≥1 Gd⁺ lesion</th>
<th>≥1 Relapse, ≥1 Gd⁺ lesion AND ≥9 T2 lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of matched patients in each treatment group</td>
<td>366</td>
<td>193</td>
<td>55</td>
<td>92</td>
</tr>
<tr>
<td><strong>ARR, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTZ</td>
<td>0.20 (0.63)</td>
<td>0.18 (0.60)</td>
<td>0.16 (0.39)</td>
<td>0.20 (0.75)</td>
</tr>
<tr>
<td>BRACE</td>
<td>0.63 (0.93)</td>
<td>0.60 (0.94)</td>
<td>0.79 (1.15)</td>
<td>0.54 (0.98)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Time to relapse, HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference BRACE</td>
<td>0.36 (0.28, 0.47)</td>
<td>0.41 (0.28, 0.89)</td>
<td>0.40 (0.21, 0.77)</td>
<td>0.40 (0.22, 0.70)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.006</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Time to discontinuation, HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference BRACE</td>
<td>0.73 (0.58, 0.93)</td>
<td>0.82 (0.60, 1.13)</td>
<td>0.61 (0.34, 1.10)</td>
<td>0.87 (0.55, 1.36)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.010</td>
<td>0.232</td>
<td>0.099</td>
<td>0.539</td>
</tr>
<tr>
<td><strong>Time to EDSS progression, HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference BRACE</td>
<td>0.97 (0.64, 1.47)</td>
<td>0.75 (0.40, 1.38)</td>
<td>0.68 (0.21, 2.14)</td>
<td>0.37 (0.13, 1.04)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.898</td>
<td>0.349</td>
<td>0.506</td>
<td>0.059</td>
</tr>
<tr>
<td><strong>AUC change in EDSS, EDSS-years, median regression (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference BRACE</td>
<td>-0.71 (-1.74, 0.32)</td>
<td>-0.06 (-2.83, 0.71)</td>
<td>0.09 (-2.66, 2.84)</td>
<td>-1.06 (-2.83, 0.71)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.174</td>
<td>0.924</td>
<td>0.949</td>
<td>0.236</td>
</tr>
</tbody>
</table>

NTZ=natalizumab; BRACE= Betaferon®/Betaseron®/Rebif®/Avonex®/Copaxone®/Extavia®; ARR=annualized relapse rate; HR=hazard ratio; CI=confidence interval; SD=standard deviation; Gd⁺=gadolinium-enhancing; EDSS=expanded disability status scale; AUC=area under the curve
Treatment persistence

During the study, 108 (29.5%) patients who initiated natalizumab ceased treatment compared with 229 (62.6%) of BRACE patients. This translated into a modelled 27% decrease in discontinuation rate (HR=0.73, 95%CI=0.58-0.93) favouring first line natalizumab (Figure 2B). Among those patients who discontinued, treatment cessation occurred earlier in the natalizumab arm than in BRACE (median [IQR] years to discontinuation: natalizumab, 1.7 [0.9-2.8]; BRACE, 2.3 [1.1-4.2]) (p(\text{signed-rank})<0.0001).

Whilst patients on first-line BRACE therapy recorded a longer mean follow-up duration the competing risks extension to the primary Cox marginal models for both relative relapse and discontinuation rates did not significantly alter the results of the primary models, suggesting that the follow-up differential observed did not significantly influence the estimated hazard ratios, either in the presence or absence of simultaneous censoring of the matched pair. In a sensitivity analysis, patients were re-matched with clinic country included as a baseline covariate. Although results were largely concordant with those demonstrated in the primary analysis, effect sizes were generally less significant, consistent with the smaller size and reduced power of these groups (Supplemental Table 1).

Disability progression

Five-hundred and thirty-four patients (natalizumab, 234; BRACE, 300) were included in the analysis of disability progression metrics. There was no significant imbalance in baseline factors between the subset treatment groups. There were no differences in 3-month confirmed disability progression events between groups (natalizumab, 16.2%; BRACE, 22.0%, \(p(McNemar)=0.095\)). There was no difference in rate of confirmed progression across either the first 12 months of treatment (HR=0.74, 95%CI=0.37-2.68, reference=BRACE) or the full treatment period (HR=0.97, 95%CI=0.64-1.47, reference=BRACE) (Table 3).
**AUC change in EDSS**

Of the 732 matched patients, 358 (natalizumab, 170; BRACE, 188) were eligible for the exploratory AUC analysis. No difference in baseline factors was observed by subset treatment group. First-line natalizumab was associated with a decrease in EDSS of a median 0.71 points per year relative to BRACE (annualised AUC change = -0.71, 95% CI -1.74, 0.32; reference=BRACE), although this difference was not significant. The proportion of patients in this subgroup recording a confirmed disability progression event over the two years was lower in the natalizumab group (17.1% vs 26.6%, p(McNemar Chi)=0.030) and there was no difference in progression rate by treatment group (HR=0.99, 95%CI=0.59-1.56; reference=BRACE).

**Subgroup analyses**

In higher pre-baseline disease activity subgroups, there were imbalance in baseline matching factors between the treatment groups after propensity matching patients in each subgroup (Table 3). Patients who initiated natalizumab as first-line therapy within all sub-groups exhibited lower on-treatment ARR relative to first-line BRACE (relative reductions, 63-80%) with the largest benefit observed in patients recording at least 2 relapses prior to initiating treatment (Table 3). Similarly, first relapse rate was lower for patients initiating natalizumab with comparable risk reductions observed in each subgroup (59%-60%) (Table 3). No differences were observed between treatment arms in treatment discontinuation, EDSS progression, or AUC in any of the subgroups (Table 3).

**Sensitivity analyses**

To test the validity of combining data from two registries we repeated the analysis sourcing first-line natalizumab initiations from within the MSBase registry. Compared with the 430 first-line natalizumab commencements available from the TOP registry, only 288 were extracted from within MSBase, of which 212 (73.6%) were able to be successfully matched on a 1:1 basis to comparable
BRACE patient. Repeating the modelling, natalizumab was again associated with both a reduction in the rate of first on-treatment relapse (HR 0.40, 95% CI 0.26, 0.63) and treatment discontinuation (HR 0.77, 95% CI 0.61, 0.99) relative to BRACE. There was no difference in disability progression (HR 0.85, 95% CI 0.22, 3.23).

Rosenbaum sensitivity analysis for the influence of unmeasured confounding
Post-estimation Rosenbaum sensitivity analyses of the outcome models estimated that a minimum 3.27, 2.90 and 2.69 fold increase in the rate of relapse, treatment discontinuation and confirmed disability progression respectively would be required for an unobserved prognostic confounder to prompt a revision of each model and reject any inference of an effect attributable to the first-line treatment arm in favour of selection effects. In the context of the hazard ratios observed, these changes represent improbably large fold increases and thus the treatment differentials observed were reasonably robust to unmeasured influences.

DISCUSSION
This is the largest reported, head-to-head comparison of on-treatment relapse occurrence, therapy persistence, and disability outcomes in patients who either initiated natalizumab or a BRACE therapy as first-line treatment. Initiating natalizumab was associated with lower ARR, lower risk of first on-treatment relapse, and lower risk of treatment discontinuation, compared to initiating a BRACE therapy. Most disability outcomes were not significantly different between the treatment groups in the primary analysis. In matched pairs completing 24 months of treatment, however, several EDSS progression measures favoured the natalizumab treatment arm, especially in patients with combined clinical and MRI disease activity and ≥9 T2 lesions.

The efficacy advantage of natalizumab compared to BRACE treatments in this study is consistent with other studies comparing these treatment options in patients who switched treatment.23,30,31 As
the population here was wholly treatment naïve, our results importantly extend these findings to potentially inform first-line treatment decisions in clinical practice. Several studies suggest that use of high efficacy treatments earlier in disease course or earlier in treatment sequence could improve long-term patient outcomes. In STRATA, an extension study for patients completing the phase III natalizumab trials, efficacy benefits associated with natalizumab (versus placebo, IFNβ, or GA) persisted after all patients were switched to natalizumab over all 6 years of observation. In the 5-year interim analysis of TOP, patients who initiated natalizumab as first-line therapy or those with lower baseline EDSS had the lowest on-treatment disease activity. Several smaller studies also suggest greater treatment response to natalizumab occurs in younger patients with lower baseline EDSS and shorter disease duration, suggesting a potential advantage for earlier natalizumab treatment initiation. Where indications permit, physicians do consider the use of natalizumab as a first-line therapy, and the current study validates this strategy.

Prior studies have suggested that not all patients are likely to benefit equally from natalizumab as a first-line therapy, with larger relative natalizumab treatment effects in patients with higher ARR in the 12 months prior to treatment initiation. Consistent with these studies, prior relapse activity appeared to influence the relative treatment effect; the relative ARR reduction seen with natalizumab treatment over BRACE therapies was observed to be highest in the patient subgroup with ≥2 relapses in the prior year (ARR was roughly 5-fold higher for BRACE) compared to groups that included patients with ≥1 relapse in the prior year (ARR was roughly 3-fold higher). Maximal relative benefits to ARR were observed in patients with more aggressive MS. Whilst this suggests relapse activity in treatment-naïve patients may enable early identification of patients who may benefit most from timely treatment intervention, efficacy benefits were apparent with natalizumab first-line treatment across all levels of disease activity. Conversely, the improvement in first-line treatment persistence in favour of natalizumab observed in the primary analysis was not replicated in any of the subgroup analyses. Whilst this may in part be secondary to under-powering associated
with the smaller subgroup sample sizes, thus may suggest discontinuation decisions may in part be influence by baseline disease severity.

Confirmed disability progression, as both a proportion of sample and as a time to event outcome, was generally not different between treatment arms. By requiring a minimum of three temporally separated EDSS assessment points, disability progression is more follow-up intensive and thus this part of the analysis is likely to be underpowered relative to either the relapse or discontinuation modelling. Matching natalizumab initiators with comparable BRACE patients on baseline similarities in disease activity favours inclusion of more benign natalizumab patients. Exclusion therefore of the most active natalizumab patients, whilst ensuring balance in these disease activity metrics across treatment arms, also biases against observation of a difference in progression by treatment group. Furthermore, higher treatment discontinuation in the BRACE group limited the total sample size available for these analyses, so that larger sample is required to determine if disability outcome differences exist between these treatments when used as first-line therapies. We were unable to adjust for the influence of JC virus antibody titre on the probability of early natalizumab discontinuation secondary to incomplete data. Even so, this suggests our observation of a significant decrease in natalizumab discontinuation rate relative to BRACE is likely to be a conservative estimate. Recent observations by Prosperini et al of an increased risk of disease worsening following natalizumab discontinuation suggests that concerns around disease reactivation may in part contribute to the observed reduction in natalizumab discontinuation rate relative to BRACE, although this was not able to be directly assessed.\textsuperscript{36}

This study is limited by non-random assignment of patients. Unlike a truly randomized design, our results may be biased from residual confounding secondary to imbalance of unobserved factors not included as a matching variable. The retrospective nature of the study means the analysis is more prone to selection and ascertainment bias, relative to a prospective design. It is, however, unlikely
that residual bias would favour the natalizumab arm, since measured variables suggest patients initiating natalizumab had, as expected, much worse disease. Whilst propensity matching cannot eliminate residual confounding secondary to unobserved influences, the Rosenbaum sensitivity analysis suggested that any confounding contributed by unmeasured influences was highly unlikely to be large enough to change these inferences. Generalisation of the efficacy advantages observed in this study could be limited by the characteristics of this population or by potential treatment indication bias that were not adjusted for in the matched datasets. The efficacy measures addressed in this study are not the only factors to consider when selecting MS treatments. Whether considering natalizumab as first-line therapy or later in treatment sequence, it is important to weigh the potential efficacy benefits of natalizumab against a patient’s individualized risk for PML when making treatment decisions. Recent observations of possible rebound disease activity following cessation of natalizumab further suggest such risk stratification may also extend to decisions around discontinuing, as well as initiating, natalizumab. A comparative analysis of safety and adverse event data was unable to be conducted secondary to insufficient data availability. A larger dataset with more complete data and longer follow-up would be required to better analyse both cessation reason and subsequent post-discontinuation outcomes. We were unable to compare MRI lesion activity outcomes by treatment group secondary to incomplete recording of post-baseline MRI. Similarly, reason for treatment discontinuation by treatment arm was only partially recorded and was thus insufficiently available to analyse. Finally, a larger dataset with longer cumulative follow-up would be required to validate the AUC analysis against hard disability end-points such as time to EDSS 6.

Our results suggest that use of natalizumab as a first-line treatment for RRMS greatly reduces relapse rate and improves medication persistence compared to the common practice of BRACE initiation. Further analyses for these first-line treatment comparisons might include cost-effectiveness and quality of life metrics, since comparative cost-effectiveness between first-line
BRACE therapies has been well-studied, and natalizumab treatment has demonstrated benefits to quality of life and patient-reported outcomes in other treatment contexts.

REFERENCES


G.3 Risk of early relapse following switch from injectables to oral agents for multiple sclerosis

**Title:** Risk of early relapse following switch from injectables to oral agents for multiple sclerosis


**Authors:** Tim Spelman¹² (MBBS), Linda Mekhael³ (PhD), Therese Burke³ (PhD), Helmut Butzkueven¹² (PhD), Suzanne Hodgkinson⁴ (PhD), Eva Havrdova⁵ (PhD), Dana Horakova⁵ (PhD), Pierre Duquette⁶ (MD), Guillermo Izquierdo⁷ (PhD), Francois Grand'Maison⁸ (PhD), Pierre Grammond⁹ (MD), Michael Barnett¹⁰ (PhD), Jeannette Lechner-Scott¹¹ (MD), Raed Alroughani¹² (MD), Maria Trojano¹³ (MD), Alessandra Lugaresi¹⁴ (MD), Franco Granella¹⁵ (MD), Eugenio Pucci¹⁶ (PhD) & Steve Vucic³ (PhD) on behalf of the MSBase Study Group

**Affiliations:**

¹ Department of Neurology, Royal Melbourne Hospital, Parkville, Australia

² Department of Medicine (RMH), The University of Melbourne, Parkville, Australia

³ Westmead Hospital, Sydney, Australia

⁴ Liverpool Hospital, Sydney, Australia

⁵ Charles University, Prague, Czech Republic

⁶ Hôpital Notre Dame, Montreal, Canada

⁷ Hospital Universitario Virgen Macarena, Sevilla, Spain

⁸ Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada

⁹ Center de réadaptation déficience physique Chaudière-Appalache, Levis, Canada

¹⁰ Brain and Mind Research Institute, Sydney, Australia

¹¹ John Hunter Hospital, Newcastle, Australia
Corresponding author:

Professor Steve Vucic

Department of Neurology, Westmead Hospital

Cnr Hawkesbury and Darcy Roads,

Westmead, NSW, 2145, Australia

Ph: +61 98456097

Email: s.vucic@neura.edu.au

Steve.vucic@sydney.edu.au

Running title: Early relapse risk post-switch to oral MS treatment

Keywords: treatment switching, multiple sclerosis, fingolimod, dimethyl fumarate, teriflunomide, relapse, progression

Word count:

Abstract: 238

Text: 2592

Figures: 2
Timothy Denis Spelman (58172)

Tables: 2

Author Disclosures:

Tim Spelman received honoraria for consultancy, funding for travel and compensation for serving on scientific advisory boards from Biogen Idec Inc; speaker honoraria from Novartis.

Linda Mekhail did not declare any competing interests.

Therese Burke received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Bayer; speaker honoraria from Merck Serono Australia, Genzyme and Biogen Idec and travel support from Biogen Idec Australia and Novartis Australia.

Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital Friends of the Neurosciences Foundation, and the University of Melbourne.

Suzanne Hodgkinson did not declare any competing interests.

Eva Havrdova received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

Dana Horakova received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

Pierre Duquette has received honoraria for organising CME events and has obtained funding to attend meetings from Biogen Idec, EMD Serono, TEVA Neuroscience, Novartis, and Genzyme, has received funding for investigator-initiated trials with Biogen Idec, EMD Serono and Novartis, and has received peer-review funding from CIHR and from the MS Society of Canada.
Guillermo Izquierdo received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

François Grand’Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.

Pierre Grammond is a Novartis, Teva-neuroscience, Biogen Idec advisory board member, consultant for Merck Serono, received payments for lectures by Merck Serono, Teva-Neuroscience and Canadian Multiple sclerosis society, and received grants for travel from Teva-Neuroscience and Novartis.

Michael Barnett has received honoraria for participation in advisory boards and travel sponsorship from Novartis, BioCSL, Genzyme and Biogen Idec

Jeannette Lechner-Scott has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen Idec, CSL, Genzyme Sanofi, Merck Serono and Novartis.

Raed Alroughani received honoraria from Biologix, Bayer, Merck Sorono, GSK and Novartis, and served on advisory board for Biologix, Novartis and Merck Sorono

Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

Alessandra Lugaresi is a Bayer Schering, Biogen Idec, Genzyme, Merck Serono Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, travel and research grants from the Associazione Italiana Sclerosi Multipla.
Franco Granella has served on scientific advisory boards for Biogen Idec, Novartis and Sanofi Aventis and has received funding for travel and speaker honoraria from Biogen Idec, Merck Serono, and Almirall.

Eugenio Pucci served on scientific advisory boards for Genzyme, Novartis and Biogen-Idec; he has received honoraria and travel grants from Sanofi Aventis, Novartis, Biogen Idec, Merck Serono, Genzyme and Teva; he has received travel grants from Associazione Marchigiana Sclerosi Multipla e altre malattie neurologiche.

Steve Vucic did not declare any competing interests.
ABSTRACT

**Background:** Early relapse outcomes in long-term stable patients switching from IFNβ/GA to oral therapy are unknown.

**Objective:** The objective of this study was to compare early relapse and progression in MS patients switching to oral therapy following a period of stable disease on interferon-beta or glatiramer acetate (IFNβ/GA), relative to a propensity-matched comparator of patients remaining on IFNβ/GA.

**Methods:** The MSBase cohort study is a global, longitudinal registry for Multiple Sclerosis. Time to first 6-month relapse in previously stable MS patients switching from platform injectables (“switchers”) to oral agents were compared with propensity matched patients remaining on IFNβ/GA (“stayers”) using a Cox Marginal Model.

**Results:** Three-hundred and ninety-six switchers were successfully matched to 396 stayers on a 1:1 basis. There was no difference in the proportion of patients recording at least 1 relapse in the first 1-6 months by treatment arm (7.3% of switchers; 6.6% stayers, p=0.675). The mean Annualised Relapse Rate (p= 0.493) and the rate of first 6-month relapse by treatment arm (HR 1.22, 95% CI 0.70, 2.11) were also comparable. There was no difference in the rate of disability progression by treatment arm (HR 1.43, 95% CI 0.63, 3.26).

**Conclusion:** This is the first study to compare early relapse switch probability in the period immediately following switch to oral treatment in a population previously stable on injectable therapy. There was no evidence of disease reactivation within the first 6 months of switching to oral therapy.
INTRODUCTION:
Switching disease-modifying treatment in multiple sclerosis is common, often triggered by breakthrough disease.\(^1\) Where disease course is stabilised on platform injectables (IFNβ/GA), poor tolerability\(^8\)\(^-\)\(^13\) or the ease of administration offered by the increasing range of oral agents may further motivate patients and/or clinicians to consider a treatment switch. Recent concerns around disease reactivation in patients switching to fingolimod\(^2\)\(^-\)\(^7\) have highlighted the paucity of data on clinical outcomes shortly after switching to oral agents, particularly in previously stable patients seeking to switch for reasons other than breakthrough disease.

In patients with satisfactory disease control on injectables, unacceptable tolerability may form the primary trigger for considering treatment switch.\(^14\) Increasing availability of oral agents for MS has expanded treatment options. Oral agents including fingolimod, dimethyl fumarate(DMF) and teriflunomide have been reported to reduce MS relapse rates by between 32% to 55% relative to placebo in clinical trials.\(^15\)\(^-\)\(^20\) A 12-month double-blind randomised trial observed a 48% decrease in relapse frequency on fingolimod relative to IFNβ-1a-IM,\(^19\) whilst a phase 3 trial of teriflunomide observed no difference in annualised relapse rate (ARR) relative to a sub-cutaneous IFNβ-1a comparator.\(^21\) A phase 3 comparison of DMF with glatiramer acetate observed a non-significant reduction in annualised relapse rate with DMF.\(^22\)

Head-to-head comparisons of clinical outcomes in patients switching to oral therapy relative to platform injectables are limited in volume. Modelling of switch data from the one-year TRANSFORMS extension trial estimated patients switching from IFNβ-1a-IM to fingolimod on average doubled their time to first relapse.\(^23\) Observational studies of fingolimod switching have further observed a significant reductions in ARR relative to other injectable agents, including an analysis by this study group which observed fingolimod switches associated with both a lower ARR
and a lower rate of confirmed disability progression when compared against switching to IFNβ/GA.\textsuperscript{23,24}

The current evidence base for clinical outcomes in a treatment switch context is largely limited to patients with active disease. Early relapse outcomes in stable patients switching from IFNβ/GA to oral therapy are unknown. The objective of this study was to compare early ARR, time to first relapse and disability progression in MS patients switching to oral therapy following a period of stable disease on IFNβ/GA, relative to a propensity-matched comparator of stable patients remaining on pre-baseline IFNβ/GA. This method has been successfully employed previously to derive product comparisons of efficacy outcomes using MSBase registry data.\textsuperscript{25-26}

**Materials & Methods**

**MSBase Registry**

All patients contributing to this study were sourced from the international MSBase Registry, a global, longitudinal, observational registry for Multiple Sclerosis. Established in 2004, the registry prospectively collects disease-related information from consenting patients attending MS treatment centers using an internet-based, physician-owned and operated system [www.msbase.org](http://www.msbase.org).\textsuperscript{27} Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing center.

**Inclusions**

Relapsing Remitting MS (RRMS) patients aged between 18 and 75 years inclusive with a baseline EDSS between 0 and 5.5 who had been stable on IFNβ/GA for a minimum of 12 months prior to baseline and then either switched to an oral agent or remained on IFNβ/GA for a minimum of 6 months post baseline were included in this analysis. Stable disease was defined as an absence of
clinical relapse events, no gadolinium–enhancing (Gd+) lesions or an increase in the number of T2 lesions on cerebral MRI and no change in EDSS score whilst on injectable DMD treatment within the 12 months pre-baseline. All cases satisfied the Poser criteria for definite MS or the McDonald criteria for MS.²⁸–³⁰

**Outcomes & Definitions**

The primary outcome was the occurrence of a relapse event within 1 to 6 months of baseline. Baseline for the switch group was the date of initiation of an oral agent. The baseline date for patients who remained on IFNβ/GA (the “stayers”) was the date of the first observed EDSS visit from 1ˢᵗ January 2010 onwards in the presence of 12 months of stable disease as defined above. A relapse was defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 hours, in the absence of concurrent illness or fever, and occurring at least 30 days after a previous attack, also previously applied in an MSBase relapse phenotype analysis.³¹ Disability progression was analysed as a secondary variable. A confirmed disability progression event was defined as ≥3-month confirmed increases of ≥0.5 points for patients with a baseline EDSS score >5.5, ≥1.0 point for those with a baseline EDSS score between 1.0 and 5.5, inclusive, and ≥1.5 points for those with a baseline EDSS score of 0. EDSS scores recorded within 30 days after a relapse onset were excluded from analysis. A switch was defined as commencing an oral DMD within a maximum of three months of discontinuing the prior agent. The date filter for the stayers was identified and applied *a priori* to ensure both switch and stay groups were contemporaneous (the 2010 start date corresponding to that period in the registry where oral DMD initiations began to be recorded with sufficient frequency).

**Statistical analyses**

Categorical variables were summarised using frequency and percentage and compared using a chi-square test. Continuous variables were summarised using mean and standard deviation (SD) or
median and inter-quartile range as appropriate, and compared using an analysis of variance or Kruskal-Wallis test as appropriate. A Bonferroni adjustment was applied to correct for multiple comparisons. Patients switching to oral therapy were propensity matched to a comparable patient remaining on IFNβ/GA using the following baseline variables: age, sex, disease duration, EDSS, number of prior DMD commencements, proportion of disease duration on treatment and number of prior DMD starts as a proportion of baseline disease duration. The propensity score was derived using a logistic regression model where switching to oral DMD formed the dependent outcome variable and the various baseline factors were incorporated into the model as explanatory covariates. Patients from each treatment arm were 1:1 matched according to similarity of propensity score using a 5-to-1 digit matching algorithm.\textsuperscript{32} Given the larger number of IFNβ/GA stayers available, many-to-one matching (2:1, 3:1 and 4:1) were also attempted. Matching success was assessed through the analysis of paired tests and standardized differences. A Wilcoxon signed-rank test and a McNemar Chi test were used to compare the post-matching balance of continuous and categorical baseline characteristics respectively between matched groups. Imbalance was defined as an absolute standardized difference equal to or greater than 0.20.\textsuperscript{33}

Analyses of time to first relapse on treatment and time to three-month confirmed disability progression were performed using a Marginal Cox model, clustered for the matched pair. Hazard proportionality was assessed via analysis of scaled Schoenfeld residuals and for all models presented in this report, hazard proportionality was satisfied. Simultaneous censoring of the matched pair, censoring each member of the pair at the earliest censor point for each outcome analysed, was used to limit the influence of informative censoring secondary to systematic differences in follow-up duration by treatment arm on event ascertainment.

Rosenbaum sensitivity analyses across all outcomes were conducted to test the sensitivity of our propensity-matched models to unobserved heterogeneity secondary to baseline characteristics that
were either not collected or incompletely observed. All analyses were undertaken using Stata version 13 (StataCorp, College Station, Texas).

RESULTS

Patient characteristics

As at the 28th January 2015, 401 “switch” patients and 3418 “stay” patients were eligible for propensity matching. Prior to matching, patients that switched to oral medications were older, had a longer disease duration, a greater number of prior DMD therapy commencements and a higher number of prior DMD initiations as a proportion of disease duration (Table 1). There was no difference in sex, baseline EDSS and the proportion of disease duration on treatment. There were no differences in the matched sample in any of these characteristics (Table 2). One-to-one matching was preferred for the primary analysis over any of the many-to-one matches secondary to the poorer quality of the secondary, tertiary and quarternary matches. Of the matched switchers, 282 (71.2%) switched to fingolimod, 64 (16.2%) to DMF and the remaining 50 (12.6%) to teriflunomide. Median (IQR) time to switching post-IFNβ/GA cessation was 14 days (0, 39). Within the matched “stayers”, 158 (39.9%) remained on IFNβ-1a-SC, 122 (30.8%) glatiramer acetate, 67 (16.9%) IFNβ-1a-IM with the remaining 49 (12.4%) on IFNβ-1b. Fingolimod switchers were younger at baseline relative to both DMF (p(rank-sum)=0.010) and teriflunomide (p(rank-sum)=0.021) had a higher baseline EDSS relative to both DMF (p(rank-sum)=0.002) and teriflunomide (p(rank-sum)<0.001). There was no difference in gender, disease duration and prior DMD exposure by oral switch product.
## Table 1: Baseline characteristics of unmatched patients

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Switch (n=401)</th>
<th>Stay (n=3418)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, median (IQR)</td>
<td>41.7 (35.1, 49.8)</td>
<td>40.1 (32.9, 48.1)</td>
<td>0.0016</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>301 (75.1)</td>
<td>2443 (71.5)</td>
<td>0.131</td>
</tr>
<tr>
<td>Disease duration, years, median (IQR)</td>
<td>9.1 (5.7, 15.2)</td>
<td>9.0 (4.7, 14.6)</td>
<td>0.0457</td>
</tr>
<tr>
<td>EDSS, median (IQR)</td>
<td>1.5 (1, 3)</td>
<td>1.5 (1, 2.5)</td>
<td>0.3286</td>
</tr>
<tr>
<td>Number of prior DMD starts – mean (SD)</td>
<td>1.7 (1.2)</td>
<td>0.5 (0.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Proportion disease duration on treatment – mean (SD)</td>
<td>0.6 (0.3)</td>
<td>0.6 (0.3)</td>
<td>0.0884</td>
</tr>
<tr>
<td>Number of prior DMD starts as a proportion of disease duration - mean (SD)</td>
<td>0.2 (0.2)</td>
<td>0.1 (0.1)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Abbreviations: EDSS = Expanded Disability Status Scale; IQR = interquartile range; SD = standard deviation; DMD = Disease-Modifying Drug
### Table 2: Baseline characteristics of propensity score-matched patients

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Switch (n=396)</th>
<th>Stay (n=396)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, median (IQR)</td>
<td>41.7 (35.0, 49.8)</td>
<td>42.9 (35.1, 51.5)</td>
<td>0.2121</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>298 (75.3)</td>
<td>296 (74.8)</td>
<td>0.8676</td>
</tr>
<tr>
<td>Disease duration, years, median (IQR)</td>
<td>10.7 (6.1, 16.7)</td>
<td>11.5 (6.3, 17.4)</td>
<td>0.1531</td>
</tr>
<tr>
<td>EDSS, median (IQR)</td>
<td>2.0 (1.3)</td>
<td>2.0 (1.3)</td>
<td>0.5290</td>
</tr>
<tr>
<td>Number of prior DMD starts – mean (SD)</td>
<td>1.7 (1.1)</td>
<td>1.7 (1.2)</td>
<td>0.2710</td>
</tr>
<tr>
<td>Proportion disease duration on treatment – mean (SD)</td>
<td>0.6 (0.3)</td>
<td>0.6 (0.3)</td>
<td>0.8991</td>
</tr>
<tr>
<td>Number of prior DMD starts as a proportion of disease duration - mean (SD)</td>
<td>0.2 (0.2)</td>
<td>0.2 (0.2)</td>
<td>0.1622</td>
</tr>
</tbody>
</table>

Abbreviations: EDSS = Expanded Disability Status Scale; IQR = interquartile range; SD = standard deviation; DMD = Disease-Modifying Drug

*All post-matching baseline characteristics had a standardized difference between -20% and +20%.
Relapse

Fifty-five (6.9%) of the 792 matched patients recorded a relapse event in the 6 months following baseline. There was no difference in the proportion of patients recording at least 1 relapse in the first 1-6 months post-baseline by treatment arm (7.3% of switchers compared with 6.6% stayers, p(McNemar)=0.675). In addition, there was no difference in the mean (SD) annualised relapse rate (ARR) within the first 6 months of baseline between the switch group (0.2 [0.7]) and the stayers (0.2 [0.6], p(signed-rank)=0.4931). There was no difference in the rate of first 6-month relapse by matched study arm (HR 1.22, 95% CI 0.70, 2.11) (Figure 1).

Figure 1 – Time to first relapse within 6 months post-baseline

Disability progression
Within the first 6 months after switching to oral therapy there were 9 (2.3%) confirmed progression events compared with 4 (1.0%) in the stayers ($p(McNemar)=0.269$). In addition, there were no differences in the rate of first disability progression events between switchers and stayers (HR 1.43, 95% CI 0.63, 3.26, reference=stayers).

**Subgroup analysis**

Amongst switchers, there was no difference in the rate of first 6-month relapse between the three oral agents (DMF adjusted HR: 1.38, 95% CI 0.56, 3.42; teriflunomide: aHR 0.58, 95% CI 0.13, 2.39; reference = fingolimod), when adjusted for baseline differences in age and EDSS. Amongst switchers the length of washout period was not associated with either early relapse (HR 0.83, 95% CI 0.62, 1.11; reference= less than 2 months) or significant disability progression (HR 1.33, 95% CI 0.68, 2.62; reference=less than 2 months).

**Sensitivity analyses**

Reapplying these models across full patient post-baseline on-treatment follow-up, there was no difference in time to any first relapse event between switchers and stayers (HR 0.87; 95% CI 0.63, 1.21, reference=stayers) (Figure 2). Similarly there was no difference in disability progression on oral therapy when compared with the stayer group (HR 1.48, 95% CI 0.79, 2.75, reference=stayers). Furthermore there was no difference in post-baseline on-treatment EDSS change between the switch group (mean (SD): 0.25 (1.21)) and stay group (mean (SD): 0.40 (1.06), $p(\text{signed-rank})=0.1799$). Mean (SD) post-baseline on-treatment follow-up was 2.1 (1.8) years. There was no difference in the rate of any first progression event between any of the three oral agents.
Figure 2 – Sensitivity analysis: Time to any first post-baseline relapse

Unmeasured confounding

Rosenbaum sensitivity analyses of our propensity-matched relapse and progression models estimated that an unobserved confounder would need to produce a minimum 1.89 and 2.07 fold increase in the rate of relapse and confirmed disability progression respectively in order to reject the inference of a treatment effect in favour of selection effects. These represent improbably large differences in the context of the observed point estimates and associated confidence intervals. This suggests that the effects of unmeasured confounding on the observed associations between treatment arm and both relapse and progression outcomes were not large enough to significantly change the inferences made.

DISCUSSION
Poor tolerability of IFNβ/GA in patients demonstrating otherwise good control of disease on treatment is an increasingly common challenge for both patients and clinicians. Sub-optimal adherence secondary to administration issues or side effect profiles of platform injectables, increases the probability of breakthrough disease, which in turn risks undoing the good work of prior therapy. The broader range of oral therapies now available has considerably expanded the options for sequencing treatments in response to breakthrough disease and tolerability issues. This is the first study to examine the real world risk of early relapse in stabilised MS patients switching to oral therapy against a propensity matched comparator arm of patients with similar duration of stable disease who remain on IFNβ/GA.

Treatment benefits observed in controlled trial settings can only be realised in clinical practice if patients are fully adherent with treatment regime. Administration convenience and patient preference for oral formulations, particularly in the setting of injection anxiety and previous experience with injection site reactions, may improve treatment adherence and thus maximise the efficacy available from oral agents. Such gains need to be balanced with potential risk of disease reactivation in the months shortly following treatment switch. The risk of early relapse following oral switch may be weighted higher by patients in our study population of long-term disease stable switchers relative to patients switching secondary to breakthrough disease. Our results suggest that early relapse in previously stable patients following a switch to oral agents was not only rare (only 7.3% of 396 switchers recorded a relapse within 1-6 months post switch), but further that there was no difference in either the proportion of patients recording an early relapse event, the ARR or time to first relapse in switchers compared to stable patients who remained on baseline IFNβ/GA therapy.

Disability progression, as measured by the EDSS, was also infrequent and no difference was observed between switchers and stayers within either the first six months post-baseline. Whilst the present study did not establish significant differences in short-term disability progression, this
analysis was only exploratory since progression, unlike relapse rates, is observed and confirmed over several time points. It is unlikely that six months would be sufficiently long enough to characterise confirmed disability progression with appropriate power. This is consistent with the experience of clinical trials using disability progression as a study end-point.

On sensitivity analysis, stable patients switching to oral treatment demonstrated no difference in either rate of first relapse and first confirmed disability progression over long-term follow-up relative to the “stayers”. The present findings suggest that disease remains quiescent over both the immediate post-switching period (1-6 months) and longer term treatment horizon, at least in this population of previously stable patients. In any event, whilst no increase in event risk was observed in our study, a potential risk of increased disease activity needs to be considered when transitioning a stable RRMS patient from an injectable to oral treatment. In addition, the side-effect profile of oral agents including, but not limited to, bradycardia, gastrointestinal symptoms, flushing, hair-thinning,10,17,20,35 needs to be considered, particularly in a population of long-term stable patients considering regimen change. Whatever the treatment decision, close monitoring of patients switching to oral agents is suggested given long-term safety profiles are less well-established relative to the platform injectables.

A limitation of this study stems from uncontrolled confounding of prognostic factors not included in the propensity match. Confounders that were either not collected or incompletely observed in the registry, such as cerebral and spinal cord MRI disease burden, were not formally adjusted. However, unobserved confounding in this case is unlikely to significantly influence the finding in the present study as indicated by the Rosenbaum sensitivity analysis. Another potential limitation of this study is that clinician documented reasons for discontinuation were not available for all switches and thus were insufficiently powered to analyse.
The present study argues against the occurrence of rebound disease within the first 6 months of switching to oral therapy in patients that were stable or injectable or infusible treatments. Switching to an oral agent, in the setting of stable disease, needs careful consideration of the longer-term relapse risk (>6 months). Further studies assessing the factors underlying the increase in longer-term disease activity need to be undertaken in order to develop more appropriate management algorithms.

Author contributions:

T Spelman conceptualized and designed the study, conducted the statistical analysis, interpreted the analysis and drafted and revised the paper

L Mehkael interpreted the analysis and revised the paper

T Burke interpreted the analysis and revised the paper

H Butzkueven interpreted the analysis and revised the paper

S Hodgkinson interpreted the analysis and revised the paper

E Havrdova interpreted the analysis and revised the paper

D Horakova interpreted the analysis and revised the paper

P Duquette interpreted the analysis and revised the paper

F Grand-Maison interpreted the analysis and revised the paper

G Izquierdo interpreted the analysis and revised the paper

P Grammond interpreted the analysis and revised the paper

M Barnett interpreted the analysis and revised the paper

J Lechner-Scott interpreted the analysis and revised the paper

R Alroughani interpreted the analysis and revised the paper

M Trojano interpreted the analysis and revised the paper

A Lugaresi interpreted the analysis and revised the paper

F Granella interpreted the analysis and revised the paper
E Pucci interpreted the analysis and revised the paper

S Vucic conceptualized and designed the study, interpreted the analysis and drafted and revised the paper

REFERENCES


G.4 Comparison of efficacy and persistence of first line fingolimod vs interferon-beta/glatiramer in the presence of prior disease activity

Title: Comparison of efficacy and persistence of first line fingolimod vs interferon-beta/glatiramer in the presence of prior disease activity (paper-format report)

AUTHORS: Tim Spelman¹,², Guillermo Izquierdo³, Raed Alroughani⁴, Ricardo Fernández Bolaños⁵, Eva Havrdova⁶, Dana Horakova⁶, Celia Oreja-Guevara⁷, Jeannette Lechner-Scott⁸, Mark Slee⁹, Michael Barnett¹⁰, Cavit Boz¹¹, Murat Terzi¹², Francois Grand'Maison¹³, Alessandra Lugaresi¹⁴, Maria Trojano¹⁵ and Helmut Butzkueven¹,², on behalf of the MSBASE Study Group

¹ Department of Neurology, Royal Melbourne Hospital, Parkville, Australia
² Department of Medicine (RMH), The University of Melbourne, Parkville, Australia
³ Hospital Universitario Virgen Macarena, Sevilla, Spain
⁴ Amiri Hospital, Kuwait City, Kuwait
⁵ Hospital Universitario Virgen de Valme, Sevilla, Spain
⁶ Charles University, Prague, Czech Republic
⁷ Hospital Universitario La Paz, Madrid, Spain
⁸ John Hunter Hospital, Newcastle, Australia
⁹ Flinders Medical Centre, Adelaide, Australia
¹⁰ Brain and Mind Research Institute, Sydney, Australia
¹¹ Karadeniz Technical University, Trabzon, Turkey
¹² Mayis University, Medical Faculty, Samsun, Turkey
¹³ Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
¹⁴ MS Center, Department of Neuroscience, Imaging and Clinical Sciences, University ‘G. d’Annunzio’, Chieti, Italy
ABSTRACT

**Objective:** To compare annualized relapse rates (ARR), time to first relapse, treatment persistence, EDSS change and disability progression in a propensity-matched, treatment-naïve MS patients with prior disease activity on first-line fingolimod relative to first-line interferonβ or glatiramer acetate (IFNβ/GA).

**Methods:** The MSBase study is a global, longitudinal, observational registry for Multiple Sclerosis. All patients included in the analysis had at least one relapse in the 12 months prior to baseline. First line fingolimod initiations were 2:1 propensity matched to first-line IFNβ/GA commencements using sex, age, country, treatment start date, disease duration, EDSS, cerebral MRI and pre-treatment relapse activity as baseline matching characteristics. Matching quality was assessed via signed-rank and McNemar Chi tests and analysis of standardized differences. Predictors of time to first relapse, treatment discontinuation and disability progression were investigated using a clustered Cox marginal model, with simultaneous censoring of the matched pair to adjust for differences in follow-up duration.

**Results:** A total of 229 first-line fingolimod patients were successfully 1:2 matched to 458 IFNβ/GA initiations. Mean (SD) on-treatment ARR on first-line fingolimod was decreased by 43% to 0.28 (0.76) relapses per year from 0.49 (0.84) on IFNβ/GA treatment (p(signed-rank)<0.0001). Fingolimod was associated with a 40% reduction in the rate of first on-treatment relapse compared with first-line IFNβ/GA (HR 0.60, 95% CI 0.39, 0.94). First line fingolimod was associated with a 49% reduction in treatment discontinuation compared with IFNβ/GA (HR 0.51, 95% CI 0.37, 0.70). There was no difference in the rate of three-month confirmed disability progression (HR 1.95; 95% CI 0.85, 4.50; reference = IFNβ/GA).
Interpretation: First-line fingolimod was associated with a reduced ARR, rate of first on-treatment relapse and rate of treatment discontinuation relative to a propensity-matched cohort of first-line IFNβ/GA initiators.
INTRODUCTION:

Multiple sclerosis (MS) is an immune-mediated, chronic inflammatory disease for which currently there is no curative therapy. Pharmaceutical treatment is aimed at managing symptoms, preventing exacerbations and stabilizing a patient’s disease course. Recent developments in the trialling and approval of oral disease-modifying drugs (DMDs) such as fingolimod, dimethyl fumarate and teriflunomide for management of relapsing forms of MS have the potential to significantly expand the therapeutic options available to both patients and clinicians. Until Fingolimod was FDA approved for first-line therapy in 2010, the options available for first-line treatment were limited to a small set of injectable platform therapies which have thus far proven only partially efficacious in clinical practice, both in terms of preventing relapse and arresting disease progression. Since 2010, fingolimod has been further approved as a first-line agent in Switzerland, Canada, Australia and has been observed to be a popular alternative to the older platform injectables, particularly in patients unable to tolerate frequent injections.

Whilst the superior efficacy of fingolimod relative to placebo has been well established through clinical trial, the evidence base for head-to-head comparisons of fingolimod relative to the platform injectables is less well established, particularly in the first-line setting. The phase 3, twelve-month TRANSFORMS trial compared fingolimod directly with intra-muscular IFNβ-1a in a double-blind, double dummy randomized controlled trial (RCT), observing a significant reduction in the ARR and MRI outcomes favouring the fingolimod arm. The subsequent extension study confirmed this advantage favouring fingolimod over a longer 24 month interval and further observed trial patients switching from IFNβ-1a to fingolimod demonstrated a lower ARR post-switch relative to their pre-switch relapse rate. Unlike the placebo-controlled FREEDOMS study, neither the original trial nor its extension demonstrated a significant difference in disability progression between trial arms.
A 2012 systematic review of the, at the time, existing suite of both head-to-head and placebo-controlled RCTs involving first-line DMDs included an indirect comparison of fingolimod against the interferon and glatiramer-based injectables.\textsuperscript{11} Using a Mixed Treatment Comparison, the meta-analysis was able to extend the previous controlled trial observation of a reduction in ARR on fingolimod relative to IM IFNβ-1a only to all the first-line platform injectable therapies in the first-line setting. However, as this was not a direct comparison, combining data from multiple trials employing different inclusion criteria, and importantly for the primary endpoint, different definitions of relapse meant these efficacy estimates and thus any inferences made based on these estimates may be limited by between-trial heterogeneity and differences in outcome ascertainment, systematic discrepancies that can be difficult, if not impossible, to adjust for within an indirect comparison.

There is even less of an evidence base for direct head-to-head comparisons of first-line DMD persistence. Compliance with injectables has previously been observed to be sub-optimal secondary to the frequent injections or regimen complexity.\textsuperscript{12-14} A phase 4 multinational, observational study of DMD adherence amongst 2648 patients on platform injectables observed injection-related reasons, including injection anxiety, injection reactions and pain, accounted for the second most commonly cited reason for non-adherence or discontinuation.\textsuperscript{15} A retrospective cohort study of 1891 MS patients found fingolimod initiators were less likely to discontinue and generally more compliant with treatment compared with injectables.\textsuperscript{16} Treatment persistence has further been associated with a reduction in MS-related inpatient hospital admissions, MS-related medical costs and improved quality of life.\textsuperscript{15-17} Despite this, there are no direct comparisons currently available comparing treatment persistence with oral fingolimod against the platform injectables in the first-line setting. Selecting one treatment agent from the range of DMDs available for first line therapy has been observed to be a function of perceived efficacy combined with safety, administration mode
and cost considerations, further underscoring the need to expand the evidence base of head-to-head product comparisons.

The objective of our study was to directly compare ARR, time to first on-treatment relapse, EDSS change, disability progression and treatment persistence in first-line fingolimod with propensity-matched first-line interferonβ or glatiramer acetate (IFNβ/GA) treatments. A propensity-based approach was chosen as it permits a direct comparison of treatment arms to be made, balanced for the differences in the distribution of known or suspected baseline and pre-baseline confounders between the fingolimod treatment arm and IFNβ/GA comparator. This method has been successfully employed previously to derive head-to-head product comparisons of efficacy and persistence outcomes using MSBase registry observational data.19,20

Materials & Methods

MSBase Registry

All patients contributing to this study were sourced from the international MSBase Registry, a global, longitudinal, observational registry for Multiple Sclerosis, an internet-based, physician-owned and operated system www.msbase.org21. Each center enters patient data either in the offline iMed® local electronic database or the online MSBase registry data entry system during routine clinic visits and intermittently uploads codified datasets to the MSBase server. Physicians record clinical information such as date of MS onset, diagnostic category, Kurtzke Expanded Disability Status Scale (EDSS) score, relapse onset dates and characteristics and cerebral MRI. Patient records are classified as complete, and thus eligible for analysis, if they comply with the minimum upload frequency and a minimum required dataset including demography (sex, birth date, MS onset date), patient assessments (visit date, KFS, EDSS), relapses (relapse date, region affected, corticosteroid treatment), para-clinical tests (test date and type including MRI, cerebrospinal fluid examination and biochemistry) and treatment (treatment name, commencement and discontinuation dates). The use
of the iMed© electronic database and minimum dataset requirements for inclusion into MSBase ensures a unified approach; cases fulfil the Poser or McDonald criteria for MS and clinical information including relapse data is collected in “real time”. Quality assurance was maintained with inbuilt data quality checking in the iMed© local record system, which applied to key dates and data in the minimum dataset. In order to ensure EDSS competency, all participating neurologists completed the Neurostatus certification (http://www.neurostatus.net) or provided evidence of prior completion of this certification. Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing center.

Inclusions

Relapsing-remitting MS (RRMS) patients initiating first-line fingolimod or IFNβ/GA treatments were included in the analysis if they persisted on first-line treatment for a minimum of three months, recorded a cerebral MRI within six months of treatment start and, for the primary analysis, a minimum of one relapse within the 12 months prior to first-line treatment initiation. Baseline age, sex, country, disease duration, EDSS, cerebral MRI lesion type, distribution and frequency and relapse activity in the 12 and 24 months prior to baseline were the minimum data required to be included in the propensity match which, by definition, required each patient to contribute a minimum two years of pre-baseline follow-up. All cases satisfied the Poser criteria for definite MS (Poser, 1983) or the McDonald criteria for MS (McDonald, 2001).

Outcomes & Definitions

The primary outcomes of this analysis were the Annualized Relapse Rate (ARR), time to first on-treatment relapse, time to treatment discontinuation and time to three-month confirmed disability progression. EDSS change from baseline and serial disability time metrics were analysed as secondary outcomes. The baseline for this analysis was defined as recorded date of first-line DMD
initiation. A relapse event is defined as occurrence of new symptoms or exacerbation of existing symptoms persisting for at least 24 hours, in the absence of concurrent illness or fever, and occurring at least 30 days after a previous attack, also previously applied in an MSBase relapse phenotype analysis. Confirmed disability progression events were defined as ≥3-month confirmed increases of ≥0.5 points for patients with a baseline EDSS score >5.5, ≥1.0 point for those with a baseline EDSS score between 1.0 and 5.5, inclusive, and ≥1.5 points for those with a baseline EDSS score of 0. EDSS scores recorded within 30 days after the onset of a relapse were excluded. A minimum of 3 visits (including baseline) at which an EDSS was formally recorded were, by definition, required to be able to assess confirmed disability progression.

**Statistical analyses**

Categorical variables were summarized using frequency and percentage. Continuous variables were assessed for significant departures from normality using a Shapiro-Wilk test of skew and summarised using mean and standard deviation (SD) or standard error (SE), or median and inter-quartile range (IQR) as appropriate. A Wilcoxon rank-sum test and chi-square test were used to compare baseline characteristics by treatment arm prior to matching. A rank-sum test was further used to compare ARR and EDSS change by treatment arm in the unmatched cohort.

Patients satisfying the inclusion criteria were eligible to be propensity matched. Patients initiating first-line fingolimod were propensity matched to two comparable patients commencing first line IFNβ/GA. The propensity score was derived on a logistic regression where the theoretical receipt of first line fingolimod formed the dependent variable defined as a function of age, sex, country, disease duration, EDSS, total relapses and steroid-treated relapses in the 12 and 24 months prior to first-line DMD initiation, number of T1 gadolinium-enhancing (Gd+) lesions and number of T2 hyperintensive lesions on baseline cerebral MRI as explanatory covariates. Month of treatment
initiation was further included as an additional covariate in the derivation of the propensity score to ensure contemporaneous treatment arms. Given the far larger pool of eligible first line IFNβ/GA available for several many-to-one matching procedures were attempted (2:1, 3:1, 4:1) with a 2:1 match ultimately selected for the primary analysis secondary to superior balance and relatively poorer quality of the tertiary and quaternary matches leading to persistant significant differences in matching covariates between treatment arms and unacceptably high standardized differences. Matching success was then assessed through the analysis of paired tests and the calculation of standardized differences for each matching covariate. A Wilcoxon signed-rank test and a McNemar Chi test were used to compare the post-matching balance of continuous and categorical baseline characteristics respectively between the fingolimod and IFNβ/GA treatment arms. Imbalance was defined as an absolute value of the standardized difference equal or greater than 0.20.²⁶,²⁷

Comparative analyses of time to first relapse on treatment, time to treatment discontinuation, and time to three-month confirmed disability progression between first line fingolimod and the IFNβ/GA comparator arms were performed using a Cox Marginal Model. Hazard proportionality was assessed via analysis of scaled Schoenfeld residuals and for all models presented in this report, hazard proportionality was satisfied. Simultaneous censoring of the matched pair, censoring each member of the pair at the earliest censor point for each outcome analysed, was used to limit the influence of informative censoring secondary to systematic differences in follow-up duration by treatment arm on event ascertainment.

A post-hoc Rosenbaum sensitivity analyses across all outcomes were conducted to test the sensitivity of our propensity-matched models to unobserved heterogeneity secondary to baseline characteristics that were either not collected or incompletely observed.²⁸-³³ All analyses were undertaken using Stata version 12 (StataCorp, College Station, Texas) and R (R Foundation for Statistical Computing, Vienna, Austria).
RESULTS

Patients

As of the date of data extraction there were 259 first-line fingolimod initiations and 11,973 first-line IFNβ/GA commencements available from the registry who both satisfied the inclusion criteria and possessed the minimum dataset of baseline and pre-baseline demography, disease activity and investigation metrics required to run the propensity match. These pre-match characteristics are summarized in Table 1, disaggregated by treatment arm. Prior to matching, first line fingolimod initiators had a greater number steroid-treated relapses in the 12 and 24 months prior to initiating first-line treatment, a greater proportion of both one or more gadolinium-enhancing lesions and 9 or more T2 hyper-intensive lesions on baseline cerebral MRI and a lower baseline EDSS relative to the first-line BRACE initiators. Furthermore significant imbalance between treatment arms, as quantified by an absolute standardized difference greater than 20%, was also observed in age, disease duration and the total number of relapses in the 12 months prior to treatment. Of the fingolimod initiators, 229 (88.4%) were closely matched on propensity score to a pair of baseline-comparable IFNβ/GA initiators on a 2:1 basis to form 229 matched triplets. Table 2 demonstrates that there were no significant differences in any of the matching characteristics. Furthermore the standardized differences for all covariates were all within +/-20%, representing good balance of these factors between treatment arms.
### Table 1: Comparison of baseline characteristics - unmatched

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>Fingolimod (n=259)</th>
<th>IFNβ/GA (n=11937)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female n (%)</td>
<td>191 (73.8)</td>
<td>8527 (71.4)</td>
<td>0.4150</td>
</tr>
<tr>
<td>Age (years) – median (IQR)</td>
<td>35.42 (27.24, 43.26)</td>
<td>34.06 (27.43, 41.55)</td>
<td>0.2161</td>
</tr>
<tr>
<td>Disease duration (years) – median (IQR)</td>
<td>2.63 (0.91, 6.64)</td>
<td>2.17 (0.70, 6.50)</td>
<td>0.0788</td>
</tr>
<tr>
<td>Baseline EDSS – median (IQR)</td>
<td>1.5 (1, 2.5)</td>
<td>2 (1.5, 3.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean(SD)</td>
<td>1.42 (0.64)</td>
<td>1.47 (0.71)</td>
<td>0.3350</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean(SD)</td>
<td>1.79 (0.95)</td>
<td>1.86 (0.98)</td>
<td>0.2424</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.71 (0.66)</td>
<td>0.63 (0.70)</td>
<td>0.0221</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.80 (0.74)</td>
<td>0.73 (0.83)</td>
<td>0.0154</td>
</tr>
<tr>
<td>1+ T1 gadolinium-enhancing lesions – n (%)</td>
<td>40 (15.4)</td>
<td>938 (7.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>9+ T2 hyperintensive lesions – n (%)</td>
<td>102 (39.4)</td>
<td>1918 (16.1)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

IQR = Inter-quartile range, EDSS = Expanded Disability Status Scale, SD – standard deviation, IFNβ = Interferon-β, GA = Glatiramer Acetate

### Table 2: Comparison of baseline characteristics in the 2:1 propensity matched patients

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>Fingolimod (n=229)</th>
<th>BRACE (n=458)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female n (%)</td>
<td>167 (72.9)</td>
<td>325 (71.0)</td>
<td>0.5900</td>
</tr>
<tr>
<td>Age (years) – median (IQR)</td>
<td>35.42 (27.93, 44.44)</td>
<td>35.42 (27.35, 43.58)</td>
<td>0.8434</td>
</tr>
<tr>
<td>Disease duration (years) – median (IQR)</td>
<td>2.43 (0.86, 6.64)</td>
<td>2.14 (0.60, 7.39)</td>
<td>0.5244</td>
</tr>
<tr>
<td>Baseline EDSS – median (IQR)</td>
<td>1.5 (1, 2.5)</td>
<td>2 (1, 3)</td>
<td>0.1268</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean(SD)</td>
<td>1.44 (0.65)</td>
<td>1.47 (0.65)</td>
<td>0.4309</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean(SD)</td>
<td>1.80 (0.96)</td>
<td>1.78 (0.86)</td>
<td>0.8837</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.73 (0.66)</td>
<td>0.75 (0.70)</td>
<td>0.9008</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.82 (0.77)</td>
<td>0.83 (0.73)</td>
<td>0.7448</td>
</tr>
<tr>
<td>1+ T1 gadolinium-enhancing lesions – n (%)</td>
<td>39 (17.0)</td>
<td>77 (16.8)</td>
<td>0.5760</td>
</tr>
<tr>
<td>9+ T2 hyperintensive lesions – n (%)</td>
<td>93 (40.6)</td>
<td>166 (36.2)</td>
<td>0.0670</td>
</tr>
</tbody>
</table>

IQR = Inter-quartile range, EDSS = Expanded Disability Status Scale, SD – standard deviation, IFNβ = Interferon-β, GA = Glatiramer Acetate
Mean (SD) follow-up on first-line treatment across the matched sample was 1.6 years (1.0) with IFNβ/GA initiators recording a longer on-treatment follow-up relative to first-line fingolimod (mean [SD]: IFNβ/GA 1.7 years (1.0); fingolimod 1.5 years (1.0); p(signed-rank)=0.0327). This follow-up differential between treatment arms was adjusted for using simultaneous censoring of the matched triplet as described in the statistical analyses section.

Relapse

Fifty-six (24.5%) of first-line fingolimod initiators recorded at least one relapse on treatment compared to 173 (37.8%) of BRACE patients. Mean on-treatment annualised relapse rate (ARR) on first-line fingolimod was decreased by 43% to 0.28 (0.76) relapses per year from 0.49 (0.84) on IFNβ/GA treatment (p(signed-rank)<0.0001). The incidence of first relapse on fingolimod was 20.2 first events per 100 person-years of follow-up (95% CI 13.8, 29.7) compared to 33.4 (95% CI 27.1, 41.3) on IFNβ/GA. On Cox marginal modelling using simultaneous censoring, fingolimod was associated with 40% reduction in the rate of first on-treatment relapse relative to the IFNβ/GA comparator (HR 0.60, 95% CI: 0.39, 0.94) (Figure 1).

Figure 1: Time to first relapse
Treatment persistence

First-line treatment discontinuation was more frequent in the IFNβ/GA arm with 201 (43.9%) of patients discontinuing over the study period compared with 46 (20.1%) of first-line fingolimod initiations (p(McNemar)<0.001). The incidence rate of IFNβ/GA discontinuation was 26.0 (95% CI 22.6, 29.8) cessations per 100-person years of follow-up compared with 13.1 (95% CI 9.8, 17.5) fingolimod discontinuations per 100-person years. This translated into a 49% reduction in the rate of treatment discontinuation in the time-to-event Cox marginal model (HR 0.51, 95% CI: 0.37, 0.70) (Figure 2). Interestingly, although fingolimod discontinuation was far less common than IFNβ/GA cessations, in the comparatively rare instances where fingolimod was ceased it tended to occur sooner into a patient's treatment course with a mean (SD) time to fingolimod discontinuation of 0.9 (0.6) years compared with 1.3 years (0.8) (p(signed-rank)=0.0034). Of the 46 fingolimod discontinuations, 42 (91.3%) recorded a reason for discontinuation, 14 cited a lack of improvement/relapse persistence/disease progression, 8 as a scheduled stop, 5 secondary to an adverse event and only 1 due to intolerance. By comparison only 138 (68.9%) of the 201 IFNβ/GA
discontinuations recorded a reason for discontinuation with lack if improvement/relapse persistence/disease progression cited as the most common reason (n=68), followed by adverse event (n=21) and lack of tolerance (n=15).

Figure 2: Time to confirmed disability progression

Disability progression

Of the 522 patients eligible for the exploratory three-month confirmed disability progression comparison, 23 (14.2%) of the eligible fingolimod patients recorded a progression event compared with 34 (9.4%) of eligible IFNβ/GA patients. The incidence of first progression event on fingolimod was 6.8 (95% CI 3.7, 12.7) events per 100 person-years of follow up compared to 3.7 (95% CI 2.1, 6.3) on first line IFNβ/GA. There was no difference in the rate of confirmed disability progression between treatment arms (HR 1.95; 95% CI 0.85, 4.50).

Rosenbaum sensitivity analysis for the influence of unmeasured confounding
Using post-hoc Rosenbaum sensitivity analyses of our propensity-matched relapse, persistence and progression models we estimated that an unobserved confounder would need to produce a minimum 2.23, 2.38 and 2.01 fold increase in the rate of relapse, discontinuation and confirmed disability progression respectively in order to reject the inference of a treatment effect in favour of selection effects. These represent improbably large differences in the context of the point estimates and associated confidence intervals observed, thus we can conclude that our observation of efficacy and persistence differentials by first-line treatment arm were reasonably robust towards the effect of unmeasured influences.

DISCUSSION

Choice of a first-line product and non-adherence or premature discontinuation of therapy secondary to disease exacerbation or poor tolerability presents a major challenge to both clinicians and patients. Our study is the first to directly compare efficacy and treatment persistence in fingolimod head-to-head against all available platform injectables in the first line setting. Our results corroborate the reductions in ARR previously observed with fingolimod in the pivotal TRANSFORMS trial and the subsequent Roskell systematic review, although the former was not limited exclusively to first line DMD commencements whilst the latter was not a direct comparison. Our study extends the current evidence base supporting an efficacy advantage attributable to fingolimod relative to IFNβ/GA by further observing a significant reduction in the rate of first-relapse on treatment. Furthermore fingolimod was associated with superior persistence of first-line therapy relative to IFNβ/GA.

The largest treatment effect observed in our study favouring first line fingolimod was the significant reduction in relapse rate across both the entire observed treatment period. As relapse on IFNβ/GA was both more frequent and on average occurred earlier in treatment, the use of simultaneous
censoring (censoring each member of the matched triplet at the earliest censor point recorded by either member of the triplet) suggests that the observed reduction in time to first on-treatment relapse favouring fingolimod is likely to be a rather conservative estimate.

Whilst oral formulations present obvious advantages for patients dissatisfied or otherwise intolerant of frequent injections, a patient’s experience of treatment-related adverse events, particularly in the early months following DMD initiation strongly influences subsequent persistence. Whilst we observed a clear treatment persistence advantage favouring fingolimod over IFNβ/GA therapies, in the comparatively rarer instances where a study patient ceased fingolimod it tended to occur significantly sooner after initiation relative to IFNβ/GA by an average of around 4-5 months, perhaps secondary to fingolimod-associated adverse events such as new varicella-zoster virus infections, bradyarrhythmias or skin cancers. This underscores the importance of patient-level risk stratification and balancing these with the perceived advantages associated with oral therapy, particularly when deciding upon the most appropriate agent to initiate in the first-line setting.

Whilst the proportion of fingolimod patients experiencing a three-month confirmed disability progression on treatment was higher than IFNβ/GA, there was no difference in the rate of progression by treatment arm. This is consistent with the existing literature which thus far has failed to observe significant reductions in progression rate attributable to fingolimod, at least relative to the platform injectables in either the first or later-line settings. That said, although our study reports outcomes across significantly longer treatment duration relative to the pivotal trials, our confirmed progression event data remains relatively sparse, with only 55 on-treatment first confirmed progression events observed. Thus this part of our analysis is likely underpowered and a larger study with greater cumulative on-treatment follow-up would be required to either corroborate no differential in progression by treatment arm or demonstrate an actual difference.
Whilst our study design of applying propensity matching to observational registry data does not constitute the same level of evidence as a randomized controlled trial, it does possess several advantages relative to the existing evidence base for treatment-naïve patients. In the first instance it permits a direct comparison to be made between fingolimod and all platform injectables explicitly in the first-line setting, whilst controlling for the influence of known confounders of efficacy and persistence outcomes. Secondly, it prospectively tracks patients over a significantly longer time-frame than the study intervals favoured in the pivotal clinical trials. Whilst the latter were limited to 12 or 24 months maximum follow-up, the mean follow-up in our study across the matched sample was 1.6 years with 25% of the matched sample contributing at least two years of follow-up, permitting us to observe and make inferences about the relative efficacy of first-line fingolimod across a longer treatment interval. This is particularly relevant for complex, temporal phenomena such as confirmed disability progression which commonly require more than the 12 months characteristic of the pre-extension study pivotal trials to first observe and then confirm. Thirdly, ARR as a primary clinical trial endpoint has been observed to have steadily fallen over recent years, possibly secondary to trial inclusion criteria favouring selection of patients with less severe baseline disease coupled with enhancements in within-trial rescue provisions. By comparison, the relapse data analysed in our study is sourced from a real-world, clinic environment and thus may constitute a realistic estimate of how efficacy differentials between first-line DMDs play out in clinical practice.

Our study has several limitations. Although the use of propensity matching ensured that the matched sample used in the primary analysis was well balanced for those baseline characteristics used in the match, it does not formally adjust for confounders of outcome that were either not collected in the registry or incompletely collected. Whilst propensity matching cannot eliminate residual confounding secondary to unobserved influences, we were able to estimate the minimum size required for an association between a hypothetical unobserved confounder and any of our primary outcome variable to significantly alter our model estimates and thus prompt a revision. As
described in the results, the Rosenbaum sensitivity analysis suggested that the there was no significant confounding of any of our outcome models secondary to unmeasured characteristics. Even so, the control of confounding using a propensity-based approach remains inferior to a randomized design. Clinician-documented reasons for discontinuation for the patients included in this study were incompletely recorded in the registry, particularly for the IFNβ/GA comparator arm and thus we lacked sufficient power to analyse these beyond providing descriptive statistics as outlined in the results section.

**FUNDING:**

The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Early Career Fellowship [1071124], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL.

**ACKNOWLEDGEMENTS:**

**MSBase study group co-investigators and contributors:** From the MS-Centrum Nijmegen, Nijmegen, The Netherlands Dr Cees Zwanikken; from Hospital S. Joao, Porto, Portugal, Maria Edite Rio; Veszprem Megyei Csolnoky Ferenc Korhaz, Veszprem, Hungary, Dr Imre Piroska; from Jewish General Hospital, Montreal, Canada, Dr Fraser Moore; from Josa Andras Hospital, Nyiregyhaza, Dr Tunde Erdelyi; The Alfred Hospital and Monash University, Melbourne, Australia, Dr Olga Skibina; from Cliniques Universitaires Saint-Luc, Brussels, Belgium, Dr Vincent Van Pesch; from Ospedali Riuniti di Salerno, Salerno, Italy, Dr Gerardo Iuliano; from Jeroen Bosch Ziekenhuis, Den Bosch, The Netherlands, Dr Erik van Munster; from FLENI, Buenos Aires, Argentina, Dr Marcela Fiol, Dr Jorge Correale and Dr Celica Ysraelit; from Department NEUROFARBA, Section of Neurosciences, University of Florence, Florence, Italy, Dr Maria Pia Amato; from Francicus Ziekenhuis, Roosendaal,
The Netherlands, Dr Leontien den Braber-Moerland; from New York University Langone Medical Center, New York, USA, Dr Joseph Herbert and Dr Iliya Kister; from Hopital Tenon, Paris, France, Dr Etienne Rouillet; from Jahn Ferenc Teaching Hospital, Budapest, Hungary, Dr Krisztian Kasa; from Central Clinical Emergency Military Hospital, Bucharest, Romania, Dr Carmen-Adella Sirbu; from the Geelong Hospital, Geelong, Australia, Dr Cameron Shaw; from HIGA Gral. San Martin, La Plata, Argentina, Dr Santiago Vetere; from the Westmead Hospital, Sydney, Australia, Dr Steve Vucic; from the Clinic of Neurology Clinical Center, Skopje, Macedonia, Dr Tatjana Petkovska-Boskova; from the Bombay Hospital Institute of Medical Sciences, Mumbai, India, Dr Bhim Singhal; from the Instituto de Neurociencias, Cordoba, Argentina, Dr Elizabeth Alejandra Bacile Bacile; from the Hospital Ecoville, Brazil, Dr Walter Oleschko Arruda; from the Centre hospitalier del’Universite de Montreal, Hopital Notre-Dame, Canada, Dr Elaine Roger and Dr Pierre Despault; from the Royal Melbourne Hospital, Australia, Dr Mark Marriott, Dr Anneke Van der Walt, Dr John King, Dr Jill Byron and Ms Lisa Morgan; from Box Hill Hospital, Monash University, Australia, Ms Jodi Haartsen; from Department of Neuroscience and Imaging, University ‘G. d’Annunzio’, Italy, Dr Giovanna De Luca, Dr Valeria Di Tommaso, Dr Daniela Travaglini, Dr Erika Pietrolongo, Dr Maria di Ioia, Dr Deborah Farina and Dr Luca Mancinelli; from Hospital Italiano, Argentina, Dr Juan Ignacio Rojas and Dr Liliana Patrucco; from Ospedale di Macerata, Italy, Dr Elisabetta Cartechini and Dr Giorgio Giuliani; from John Hunter Hospital, Australia, Dr David Williams and Dr Lisa Dark; from Buenos Aires, Argentina, Dr Aldo Savino; and from Sheba Medical Center, Tel Hashomer, Israel, Dr Joab Chapman; from Assaf Harofeh Medical Center, Beer-Yaakov, Israel, Dr Shlomo Flechter; from Hospital Italiano, Buenos Aires, Argentina, Dr Edgardo Cristiano; from Centro Internacional de Restauracion Neurologica, Havana, Cuba. Dr Jose Antonio Cabrera-Gomez; from INEBA, Buenos Aires, Argentina, Dr Maria Laura Saladino; from Hospital Fernandez, Buenos Aires, Argentina, Dr Norma Deri; from Craigavon Area Hospital, Portadown, UK, Dr Orla Gray; from St Vincent’s Hospital, Melbourne, Australia; Dr Mark Paine; and from Mater Dei Hospital, Malta; Dr Norbert Vella; Mr Samir Méchati, Mr Eric Bianchi, Mr Alexandru Bulla and Mr Matthieu Corageoud.
Author Disclosures:

Tim Spelman received compensation for serving on scientific advisory boards, honoraria for consultancy and funding for travel from Biogen Inc; speaker honoraria from Novartis.

Guillermo Izquierdo received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

Raed Alroughani received honoraria from Biologix, Biogen, Bayer, Genzyme, Genpahrm, Merck Sorono, GSK and Novartis, and served on advisory board for Bayer, Biologix, Biogen, Genzyme, Genpharm, Novartis and Merck Sorono.

Ricardo Fernández Bolaños did not declare any competing interests

Eva Havrdova received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

Dana Horakova received speaker honoraria and consultant fees from Biogen Idec, Merck Serono, Novartis, Genzyme and Teva, as well as support for research activities from Biogen Idec and Merck Serono.

Jeannette Lechner-Scott has accepted travel compensation from Novartis, Biogen and Merck Serono. Her institution receives the honoraria for talks and advisory board commitment and also clinic support from Bayer Health Care, Biogen, CSL, Genzyme Sanofi, Merck Serono, Novartis and TEVA.

Mark Slee did not declare any competing interests

Michael Barnett has received honoraria for participation in advisory boards and travel sponsorship from Novartis, BioCSL, Genzyme and Biogen Idec

Cavit Boz received conference travel support from Biogen Idec, Novartis, Bayer-Schering, Merck-Serono and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.
Murat Terzi received travel grants from Merck Serono, Novartis, Bayer-Schering and Teva; has participated in clinical trials by Sanofi Aventis, Roche and Novartis.

Francois Grand-Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.

Alessandra Lugaresi was a Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novartis Advisory Board Member. She received travel grants and honoraria from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, research grants from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi and Teva, travel and research grants from the Associazione Italiana Sclerosi Multipla.

Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital.

Author Contributions:

Tim Spelman conceptualized and designed the study, conducted the statistical analysis, interpreted the analysis and drafted and revised the paper

Guillermo Izquierdo interpreted the analysis and revised the paper

Raed Alroughani interpreted the analysis and revised the paper

Ricardo Fernández Bolaños interpreted the analysis and revised the paper

Eva Havrdova interpreted the analysis and revised the paper
Timothy Denis Spelman (58172)

Dana Horakova interpreted the analysis and revised the paper

Jeannette Lechner-Scott interpreted the analysis and revised the paper

Mark Slee interpreted the analysis and revised the paper

Michael Barnett interpreted the analysis and revised the paper

Cavit Boz interpreted the analysis and revised the paper

Murat Terzi interpreted the analysis and revised the paper

Francois Grand-Maison interpreted the analysis and revised the paper

Alessandra Lugaresi interpreted the analysis and revised the paper Multipla.

Maria Trojano interpreted the analysis and revised the paper

Helmut Butzkueven conceptualized and designed the study, interpreted the analysis and revised the paper

REFERENCES:


G.5 Comparison of Switching to Natalizumab versus Remaining on Interferon-Beta or Glatiramer Acetate after On-Treatment MS Relapse Using Propensity-Matched Registry Data

Title: Comparison of Switching to Natalizumab versus Remaining on Interferon-Beta or Glatiramer Acetate after On-Treatment MS Relapse Using Propensity-Matched Registry Data (paper-format report)

AUTHORS: Tim Spelman,1 Fabio Pellegrini,2 Annie Zhang,2 Maria Trojano,3 Heinz Wiendl,4 Ludwig Kappos,5 Robert Hyde,2 Shibeshih Belachew,2 Freek Verheul,6 Francois Grand-Maison,7 Guillermo Izquierdo,8 Helmut Butzkueven,1 on behalf of the MSCOMET (an MSBase Substudy) Investigators and the TOP Investigators

1Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia
2Biogen Inc., Weston, MA
3Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy
4Department of Neurology, University of Münster, Münster, Germany
5Departments of Neurology and Biomedicine, University Hospital Basel, Basel, Switzerland
6Groene Hart Ziekenhuis, Gouda, The Netherlands
7Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
8Hospital Universitario Virgen Macarena, Sevilla, Spain

Corresponding Author: Tim Spelman

Postal Address:
Melbourne Brain Centre, Royal Melbourne Hospital,
Grattan St, Parkville, Victoria, Australia 3050
Keywords: natalizumab, interferon-beta, glatiramer acetate, propensity-matching, relapse, discontinuation, treatment switching, multiple sclerosis, disability progression
Abstract:

Background: Switching immunomodulatory therapy is a potentially useful treatment strategy in patients with Multiple Sclerosis (MS) failing first line interferonβ (IFNβ) or glatiramer acetate (GA) therapy. Data for informing therapy switching is limited due to the lack of head-to-head prospective, randomized controlled trials. Outcomes in patients switching to natalizumab compared with remaining on IFNβ/GA after an on-treatment relapse are not known. This is the largest to date multi-centre, international head-to-head comparison of on-treatment relapse and treatment persistence outcomes in a cohort of patients switching to natalizumab compared with matched patients persisting on IFNβ/GA following a relapse.

Methods: Natalizumab switches sourced from Tysabri Observation Programme (TOP) registry were propensity matched to comparable IFNβ/GLA treatments sourced from the MSBase registry using a range of baseline demography and disease activity characteristics. Time to first relapse, treatment discontinuation and confirmed disability progression by treatment arm were compared using a Cox marginal model.

Results: Switching to natalizumab was associated with a 64% decrease (HR 0.36, 95% CI 0.30, 0.44) in the rate of on-treatment relapse, a 44% decrease in the rate of treatment discontinuation (HR 0.56, 95% CI 0.50, 0.63) and a 39% reduction in the rate of confirmed disability progression (HR 0.61, 95 CI: 0.36, 0.51) relative to persisting on IFNβ/GA. Natalizumab was also associated with a greater mean (SD) decrease in on-treatment EDSS (-0.04 (0.98) vs 0.06 (1.08); p=0.0159).

Conclusions: Following treatment failure on first line IFNβ/GA, switching to natalizumab may be associated with superior disease stabilisation and treatment persistence compared with persisting on IFNβ/GA therapy.
**Introduction:**

Outcomes in patients switching to natalizumab compared with remaining on IFNβ/GA after an on-treatment relapse are not known. Switching immunomodulatory therapy is a potentially useful treatment strategy in patients with MS failing first line IFNβ/GA therapy. Up to 30% of patients in the early years of treatment demonstrates suboptimal therapeutic response; however no universal guidelines for informing switch choices for patients with breakthrough disease currently exist. Data for informing such choices is limited due to the paucity of relevant, head-to-head prospective, randomized controlled trials.

Although Disease Modifying Drugs (DMDs) are usually approved on the basis of placebo-controlled randomized studies, DMD switch decisions are also not usually supported by trial evidence. Prospective MS outcome datasets acquired in the observational setting in registries or phase 4 studies represent another opportunity to generate efficacy comparisons between DMDs. Observational studies are, by definition, non-randomized, with treatments chosen as a result of patient and physician preference and the current regulatory environment. Therefore, comparison groups may vary widely from each other with regard to their pre-treatment characteristics, leading to selection bias and confounding. Such imbalance can fatally confound any demonstrated difference in efficacy by treatment arm, limiting the ability of the researcher to be able to attribute such difference to the treatment of interest and potentially compromising subsequent treatment decisions in the clinical setting. Propensity score matching is a statistical technique for correcting for covariate imbalance in non-randomly selected cohorts. Propensity based methodologies, with their ability to balance the distribution of known or suspected confounders of treatment efficacy across study arms, can reduce selection bias, and have been shown to closely approximate estimates of treatment effect derived from randomized trials, whilst preserving the real-world characteristics of observational registry data. Propensity score techniques have recently been
successfully used to correct for imbalance in baseline MS disease characteristics in comparisons of IFNβ treatment persistence and disease outcomes using MS observational registry data.\textsuperscript{10-12}

Natalizumab is a recommended treatment in patients with inadequate response to IFNβ/GA therapy or otherwise aggressive MS,\textsuperscript{13,14} and has been demonstrated observational studies to increase the proportion of these first-line failure patients who remained relapse-free, disability progression-free and MRI activity free post-switch and decrease the overall relapse rate compared to within-IFNβ or GA formulation switches.\textsuperscript{15,16} Although natalizumab has been observed in a clinical trial setting to reduce risk of relapse and disability progression, at least when compared to placebo,\textsuperscript{17} there are no head-to-head clinical trials comparing efficacy outcomes between NAT switches and IFNβ/GLA persistence following prior IFNβ/GA failure.

Our primary hypothesis for this analysis is that switching to natalizumab following at least one relapse on prior IFNβ/GLA therapy is associated with both a decreased risk of subsequent relapse and disability progression and an increased treatment persistence compared with propensity-matched patients remaining on IFNβ/GLA therapy.

**MATERIALS & METHODS**

**MSBase Registry**

The MSBase Registry is an international online database accumulator that was established in 2004 and collects disease related information from consenting patients attending Multiple Sclerosis (MS) clinics. The registry is a collaborative research group that prospectively collects outcomes data from MS treatment centres using an internet-based, physician owned and operated system [www.msbase.org].\textsuperscript{18} Each center enters patient data in the offline iMed© local electronic database during routine clinic visits and intermittently uploads anonymized datasets to the MSBase server. Physicians record clinical information such as date of MS onset, Kurtzke Expanded Disability Status...
Score (EDSS), relapse characteristics, MRI and other investigations and diagnostic criteria used. Records are classified as complete and eligible for analyses if they meet a minimum required set of data. Quality of the EDSS assessment was assured by the requirement of online Neurostatus certification at each of the participating centres. The MSBase registry was approved by the Melbourne Health Human Research Ethics Committee, and by the local ethics committees in all participating centres (or exemptions granted, according to applicable local laws and regulations). Informed consent from all patients according to local laws is required for participation in MSBase and the project holds Human Research Ethics Committee approval or exemption at each contributing center.

**MS Study Comparing Efficacy of Treatments (MSCOMET)**

The interferon-beta/glatiramer acetate arm of this study was sourced from the MS Study Comparing Efficacy of Treatments (MSCOMET). MSCOMET is a wholly prospective, longitudinal sub-study of the larger MSBase registry established to assess the real-world clinical effectiveness of treatment with interferon-beta or glatiramer acetate on relapse activity and disability progression, in addition to characterising predictors of treatment discontinuation and switching in clinical practice. As of the date compilation date (26th June 2015), the MSCOMET sub-study was tracking n=1000 patients from 28 clinics across 14 countries.

**Tysabri Observational Program (TOP)**

Subjects for the natalizumab treatment arm of this study were sourced from the Tysabri Observational Program (TOP). TOP is an observational, longitudinal registry designed to monitor the long-term safety profile of natalizumab, in addition to enabling analysis and assessment of on-treatment disease activity and disability progression. In TOP, a clinical relapse is also defined as new or recurrent neurological symptoms, not associated with fever, lasting for ≥24 hours and followed by a period of 30 days of stability or improvement. Study endpoints are assessed uniformly...
across sites. To assure standardised examinations and consistent definitions for the EDSS Functional System (FS) scores, participating physicians are provided a copy of the interactive Neurostatus Training DVD-ROM, and Neurostatus certification is highly recommended. Investigators not previously certified are offered the same online certification (http://www.neurostatus.net) available to MSBase investigators. To reduce the risk of entry error with EDSS score reporting, electronic case report forms (CRFs) were designed to automatically generate queries for data inconsistencies, including data that were out of range or otherwise invalid. As at the time of extraction there were 5691 patients commencing NAT across 16 countries. The TOP study design is in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines, and all enrolled patients provided written informed consent.

**Inclusions**

All patients included in the analysis recorded at least one relapse within the 12 months prior to baseline. For the IFNβ/GLA treatment arm, baseline was defined as the date of the first on-treatment visit within a patient’s prospective follow-up at which a full EDSS was recorded, preceded by at least 1 relapse in 12 months prior in the presence of at least 9 of these 12 months of prior IFNβ/GA therapy. Baseline for the natalizumab switch arm was defined as the date of natalizumab commencement following exposure to IFNβ/GA for at least 9 of the 12 months pre-baseline which was also associated with at least 1 relapse event within this 12 month period.

**Outcomes**

Our primary outcomes were annualized relapse rate (ARR), time to first relapse and time to treatment discontinuation. Secondary exploratory outcomes included change in on-treatment Expanded Disability Status Scale (EDSS) and three-month confirmed disability progression. Disability progression was defined as a minimum one and a half-point increase in EDSS score above a baseline EDSS of zero, a minimum one-point increase above a baseline of between 1 and 5.5 and a half-point
increase for baseline EDSS of 6 and above. To qualify as a confirmed disability progression event, these increases were required to be sustained at repeat assessment at least 3 months post the initial progression event. EDSS scores recorded during a relapse were excluded to avoid potential bias introduced by inflammatory exacerbations. Only patients who recorded a minimum of 3 on-treatment EDSS scores from baseline onwards (the minimum required to demonstrate a three-month confirmed disability progression event) were included in this analysis.

**Statistical Analyses**

Data from both registries were compiled and combined on 26th June 2015. Categorical variables were summarized using frequency and percentage. Continuous variables were assessed for significant departures from normality using a Shapiro-Wilk test of skew and summarised using mean and standard deviation (SD) or median and inter-quartile range (IQR) as appropriate. All baseline covariates common to and available from both registries were used to derive the propensity score, including age, sex, disease duration, baseline EDSS, number of pre-baseline treatment initiations, treatment duration as a proportion of disease duration, number of prior DMD initiations as a proportion of disease duration, total number of all relapses and steroid-treated relapses in the 12 and 24 months preceding baseline and country. A Wilcoxon rank-sum and a chi-square test were used to compare unmatched baseline characteristics by treatment arm as appropriate.

Propensity scores were calculated for each individual patient and represent the probability that a patient from either registry would have theoretically received the treatment of interest (natalizumab) based wholly on pre-treatment baseline characteristics. This propensity score was derived from a logistic regression model, in which receipt of natalizumab was the dependent outcome variable and the pre-treatment characteristics formed the explanatory independent variables. A single patient from the IFNβ/GA stay arm was then matched to two comparable natalizumab switches based on similarity of propensity score. Success of matching was assessed.
using both paired tests and the derivation of standardized differences. A Wilcoxon signed-rank and McNemar test were used to compare baseline characteristics in the matched data for continuous variables and proportions respectively. Standardized differences were calculated for both pre- and post-matching comparisons. Comparative analysis of time to first relapse on treatment, treatment discontinuation and three month confirmed disability progression between the matched natalizumab and IFNβ/GA groups were performed using a Cox Marginal Model. Simultaneous censoring of the matched triplet, where all three members are censored at the earliest censor point recorded by any one member of the triplet, was used to adjust the analyses for systematic differences in follow-up between treatment arms. Hazard proportionality was assessed via analysis of scaled Schoenfeld residuals. A Rosenbaum sensitivity analysis was used to test the relapse, discontinuation and progression models for the potential influence of unobserved confounding. Annualised relapse rate (ARR) was compared using a signed-rank test. Serial disability/time AUC change comparisons were performed to estimate patients’ cumulative on treatment disability experience. EDSS scores assessed within a 24-month interval from baseline were initially plotted, and AUC calculated as previously described in our published reports. All analyses were conducted in Stata version 13 (StataCorp, College Station, TX) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Of the 925 patients IFNβ/GLA stayers and 5584 switchers satisfying the inclusions criteria, 873 (94.4%) of stayers were propensity matched on a 2:1 basis to 1746 natalizumab switch patients. Prior to matching, patients switching from IFNβ/GA to natalizumab following relapse were significantly younger at baseline, a shorter disease duration, higher EDSS, a smaller proportion of disease duration on treatment and higher pre-baseline relapse activity relative to patient remaining on IFNβ/GA post-relapse (Table 1).
### Table 1: Comparison of baseline characteristics by unmatched treatment groups

<table>
<thead>
<tr>
<th>Baseline factor</th>
<th>MSCOMET (n=925)</th>
<th>TOP (n=5584)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex - n (%)</td>
<td>665 (71.2)</td>
<td>4030 (72.2)</td>
<td>0.861</td>
</tr>
<tr>
<td>Age (years) - median (IQR)</td>
<td>38.6 (31.6, 46.6)</td>
<td>37 (30, 44)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>EDSS - median (IQR)</td>
<td>2.5 (1.5, 4)</td>
<td>3.5 (2, 4.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Disease duration (years) - median (IQR)</td>
<td>8.3 (4.3, 13.4)</td>
<td>7.1 (3.3, 12.3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Proportion of disease duration on treatment - mean (SD)</td>
<td>0.6 (0.3)</td>
<td>0.5 (0.3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Number of DMD treatment starts - mean (SD)</td>
<td>1.5 (0.8)</td>
<td>1.6 (1.1)</td>
<td>0.0074</td>
</tr>
<tr>
<td>Number of DMD treatment starts / disease duration - mean (SD)</td>
<td>0.3 (0.2)</td>
<td>0.5 (0.3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean(SD)</td>
<td>1.2 (0.6)</td>
<td>2.0 (1.0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean(SD)</td>
<td>1.7 (1.0)</td>
<td>2.8 (1.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.6 (0.7)</td>
<td>1.6 (1.1)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.9 (1.0)</td>
<td>2.3 (1.5)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

In contrast, no significant differences in any of these baseline characteristics by treatment arm were demonstrated in the propensity matched sample (Table 2). Mean post-baseline, on treatment follow-up was longer in the natalizumab switch arm (2.49 years (SD 1.55)) compared to the IFNb/GA stayers (mean 1.93 years (SD 1.39)) (p(signed-rank)<0.0001). This follow-up differential was managed in the modelling analysis through application of simultaneous censoring of the matched triplet as described in the statistical analyses section above.
Table 2: Comparison of baseline characteristics by 2:1 propensity matched treatment groups

<table>
<thead>
<tr>
<th>Baseline factor</th>
<th>MSCOMET (n=873)</th>
<th>TOP (n=1746)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex - n (%)</td>
<td>628 (71.9)</td>
<td>1251 (71.7)</td>
<td>0.8780</td>
</tr>
<tr>
<td>Age (years) - median (IQR)</td>
<td>38.5 (31.6, 46.5)</td>
<td>39 (32, 46)</td>
<td>0.8334</td>
</tr>
<tr>
<td>EDSS - median (IQR)</td>
<td>2.5 (1.5, 4)</td>
<td>2.5 (1.5, 4)</td>
<td>0.1009</td>
</tr>
<tr>
<td>Disease duration (years) - median (IQR)</td>
<td>8.4 (4.2, 13.4)</td>
<td>8.0 (4.2, 13.2)</td>
<td>0.4156</td>
</tr>
<tr>
<td>Proportion of disease duration on treatment - mean (SD)</td>
<td>0.6 (0.3)</td>
<td>0.6 (0.3)</td>
<td>0.2150</td>
</tr>
<tr>
<td>Number of DMD treatment starts - mean (SD)</td>
<td>1.5 (0.8)</td>
<td>1.5 (0.9)</td>
<td>0.4018</td>
</tr>
<tr>
<td>Number of DMD treatment starts / disease duration - mean (SD)</td>
<td>0.5 (0.4)</td>
<td>0.5 (0.3)</td>
<td>0.0990</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean(SD)</td>
<td>1.2 (0.6)</td>
<td>1.3 (0.6)</td>
<td>0.4294</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean(SD)</td>
<td>1.7 (1.0)</td>
<td>1.8 (1.0)</td>
<td>0.1467</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.7 (0.7)</td>
<td>0.7 (0.6)</td>
<td>0.2941</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>1.0 (1.0)</td>
<td>1.1 (1.0)</td>
<td>0.0531</td>
</tr>
</tbody>
</table>

**Table 2: Comparison of baseline characteristics by 2:1 propensity matched treatment groups**

*Time to first relapse*

Within the propensity matched sample, 430 (24.6%) of natalizumab switchers recorded at least one on-treatment relapse post baseline, compared to 420 (48.1%) in the comparator IFNβ/GA stayer arm. Mean (SD) annualised relapse rate (ARR) was significantly lower in natalizumab switchers (mean 0.33, SD 0.95) relative to IFNβ/GA stayers (mean 0.47, SD 0.91) ($p$ (signed-rank)<0.001). Mean (SD) time to first relapse on-treatment was 14.5 months (13.0) on natalizumab compared with 11.6 months (9.9) on IFNβ/GLA. On Cox marginal modelling, switching to natalizumab was associated with a 64% reduction in the rate of on-treatment relapse compared with IFNβ/GLA (HR 0.36, 95% CI 0.30, 0.44) (Figure 1).
Figure 1: Time to first relapse (simultaneous censoring)

Time to treatment discontinuation

There were 747 (42.8%) natalizumab discontinuations post baseline compared with 515 (59.0%) in the IFNβ/GA. Mean (SD) time from baseline to discontinuation was shorter in the IFNβ/GA stayers (mean 1.32 years, SD 1.06) compared with the natalizumab switch group (mean 1.95 years, SD 1.25). The incidence rate of natalizumab cessation was 15.5 discontinuations per 100 person-years (95% CI 13.7, 17.6) compared with 37.5 (95% CI 33.5, 42.0) in patients persisting on IFNβ/GLA therapy.

Natalizumab was further associated with a 44% reduction in the rate of treatment discontinuation compared to the IFNβ/GA arm (HR 0.56, 95% CI 0.50, 0.63) (Figure 2).
Figure 2: Time to treatment discontinuation

**EDSS change and confirmed disability progression**

Switching to natalizumab was associated with a greater decrease in on-treatment EDSS change with a mean (SD) change of -0.04 (0.98) EDSS points compared with 0.06 (1.08) in the IFNβ/GA arm (p(signed rank)=0.0159). Of the original 873 IFNβ/GA patients and 1746 matched natalizumab patients, 662 and 1308 IFNβ/GA and natalizumab patients were eligible for the three-month disability progression analysis. Of these a higher proportion of IFNβ/GA stayers (n=120, 18.1%) recorded a confirmed progression event compared with n=216 (16.5%) of the natalizumab switchers, however this difference was no significant (p=0.369) when tested as simple proportions (i.e. irrespective of the timing of these events). However when time to progression event was modelled using simultaneous censoring of the matched triplets, natalizumab was associated with a 39% decrease in the rate of confirmed progression relative to IFNβ/GA (HR 0.61, 95 CI: 0.36, 0.51) (Figure 3).
Serial disability / time AUC analysis

Patients switching to natalizumab had significantly less total disability burden, as measured by standardized 24-month AUC values, compared to patients who remained on IFNβ/GA. The mean cumulative AUC was decreased by 0.41 EDSS-years in the natalizumab treatment arm compared to the BRACE treatment arm (natalizumab: mean [SD], -2.94 [1.78] EDSS-years; BRACE: -2.53 [1.68] EDSS-years; p < 0.001). On quantile regression, median standardized 24-month AUC was decreased by 1.18 EDSS-years (95% CI -1.95, 0.04) (p = 0.089) in patients who switched to natalizumab compared to those who switched to BRACE treatments, although this difference was not significant.

Rosenbaum sensitivity analysis
Post-estimation Rosenbaum sensitivity analyses of the Cox marginal models estimated that a minimum 2.89, 2.56 and 1.98 fold increase in the rate of relapse, treatment discontinuation and confirmed disability progression respectively would be required for an unobserved confounder to prompt a revision of each model and reject any inference of a treatment effect in favour of selection effects. In the context of the magnitude of the hazard ratio point estimates and associated confidence intervals observed, these required minimum changes represent improbably large fold increases and thus the treatment differentials observed for each of these three end-points were reasonably robust to unmeasured influences.

Discussion:

This is the largest to date head-to-head comparison of on-treatment relapse and treatment persistence outcomes in a cohort of patients switching to natalizumab compared with matched patients persisting on IFNβ/GLA following a relapse. This study suggests that switching to natalizumab following relapse on IFNβ or GLA may improve clinical outcomes and further treatment continuation of MS patients failing first-line IFNβ/GLA compared with persisting on platform therapies. Whilst first-line treatment initiation decisions in clinical practice are well supported by placebo controlled RCT data, subsequent alterations to regimen triggered by treatment failure are less well corroborated with an evidentiary base that can be accessed by clinicians and patients when negotiating treatment change decisions. Propensity matching, when supplemented with appropriate evaluation and sensitivity analyses, is a potentially powerful tool for studying comparative treatment efficacy in real-world settings, where randomised comparisons are inadequate, unavailable or otherwise prohibitive.

Treatment optimisation for those patients failing or achieving sub-optimal response to first-line IFNβ/GLA therapy is associated with superior long-term outcomes in terms of disease activity and
progression. Persisting on or switching between the various platform therapies is common, even in the presence of evidence that MS patients on platform therapy may benefit from switching from IFNβ/GA to natalizumab.\textsuperscript{21} Consistent with our results, Bonafede and colleagues, in a retrospective, observational cohort study of 6181 MS patients, observed greater treatment persistence on natalizumab compared with the platform therapies. Arguably related to comparable treatment benefits as demonstrated in our analysis, natalizumab initiation has been demonstrated in some settings to correlate with significant reductions in health costs, specifically those related to reductions in corticosteroid use and the frequency and length of MS-related inpatient stays.\textsuperscript{22} Our results further corroborate with those of Halpern et al who observed greater adherence and treatment persistence in patients on second-line natalizumab compared to an IFNβ-1a and -1b cohort.\textsuperscript{23}

Remodelling the primary outcomes of time to first relapse and treatment discontinuation on the unmatched data demonstrated a similar, albeit smaller, treatment benefit associated with natalizumab. This suggests that an unmatched methodology, even when fully adjusted for in a multivariable model, may systematically underestimate the true benefit attributable to natalizumab, a phenomenon that corroborates with our previous analyses of natalizumab switch efficacy using the same registry data sources as employed herein.\textsuperscript{24} This further suggests that a propensity-based approach, with its superior control of confounding through pseudo-randomisation, can potentially deliver estimates of natalizumab efficacy that better approximate those observed in the real-world clinical setting. Our study also suggests a future role for propensity matching as a robust and reliable means for collating data from separately maintained MS registries to take advantage of the explanatory power conferred by very large samples.

Supplementing traditional analyses of on-treatment relapse activity, disability progression has increasingly been used as a primary endpoint in comparative analyses of treatment efficacy.\textsuperscript{17,25,26}
Our analyses, albeit limited secondary to probable under-powering associated with the comparatively small number of progression events observed, suggests patients who switch from IFNβ/GLA based therapies to natalizumab may be less likely to experience a disability progression event compared with patients who persist on IFNβ/GLA therapy, although clearly a larger sample with greater cumulative follow-up would be required to help distinguish between a potentially underpowered protective benefit attributable to natalizumab and a genuine non-association. This is consistent with Kallweit et al who likewise observed a suggestion of a reduction in EDSS, albeit non-significant, after 1 year of treatment with natalizumab in a small observational cohort study. A larger study with greater cumulative follow-up would be required to better test this suggestion of an association.

Whilst our analysis is limited by non-random assignment, the favourable Rosenbaum sensitivity analysis suggests the propensity matched sample used in the modelling we reasonably robust to the potential influence of any unmeasured confounders no included in the derivation of the propensity score balance metric, and can thus be considered a reasonably unbiased measure of treatment effect. However, even with these propensity based adjustments our study cannot claim the same degree of confounder balance as would be expected under a truly randomised design. Even so, given natalizumab patients generally had more severe disease prior to matching (Table 1) it is probable that any residual bias would favour the IFNβ/GA stayer arm over the natalizumab switch arm. Further as natalizumab switchers were more likely to be censored early under simultaneous censoring given the longer follow-up in the natalizumab switch arm, it is also likely that the effect sizes observed represent conservative estimates of the true treatment effects. Further limiting this study, insufficient safety and adverse event data was available to conduct a comparative analysis by treatment arm. We were also unable to compare MRI lesion activity outcomes by treatment group secondary to incomplete recording of post-baseline MRI. Similarly, reason for treatment
discontinuation by treatment arm was only partially recorded and was thus insufficiently available to analyse or disaggregate further.

**Funding**

The work was supported by the NHMRC Career Development Award (Clinical) to HB [ID628856], NHMRC Project Grant [1032484], NHMRC Center for Research Excellence [Grant ID 1001216] and the MSBase Foundation. The MSBase Foundation is a not-for-profit organization that receives support from Merck Serono, Biogen Idec, Novartis Pharma, Bayer-Schering, Sanofi-Aventis and BioCSL.

**Author contributions:**

**Tim Spelman** conceptualized and designed the study, programmed and conducted the statistical analysis, interpreted the analysis and drafted and revised the paper

**Fabio Pelligrini** conceptualized and designed the study and revised the paper

**Annie Zhang** conceptualized and designed the study and revised the paper

**Maria Trojano** interpreted the analysis and revised the paper

**Heinz Wiendl** interpreted the analysis and revised the paper

**Ludwig Kappos** interpreted the analysis and revised the paper

**Robert Hyde** interpreted the analysis and revised the paper

**Shibeshih Belachew** conceptualized and designed the study, interpreted the analysis and revised the paper

**Freek Verheul** interpreted the analysis and revised the paper

**Francois Grand-Maison** interpreted the analysis and revised the paper

**Guillermo Izquierdo** interpreted the analysis and revised the paper

**Helmut Butzkueven** conceptualized and designed the study and revised the paper
Author disclosures:

Tim Spelman received compensation for serving on scientific advisory boards, honoraria for consultancy and funding for travel from Biogen Inc; speaker honoraria from Novartis.

Fabio Pellegrini is an employee of Biogen.

Annie Zhang is an employee of Biogen.

Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

Heinz Wiendl received compensation for serving on scientific advisory boards for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi; speaker honoraria and travel support from Bayer Schering AG, Bayer Vital GmbH, Biogen Idec, CSL Behring, Fresenius Medical Care, Genzyme, GlaxoSmithKline, GW, Merck Serono, Novartis, and Sanofi; compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi; research support from Bayer Vital GmbH, Biogen Idec, Merck Serono, Novartis, Sanofi Germany, and Sanofi US.

Ludwig Kappos received research support from Acorda, Actelion, Allozyne, BaroFold, Bayer HealthCare, Bayer Schering, Bayhill Therapeutics, Biogen Idec, Elan, European Union, Genmab, Gianni Rubatto Foundation, GlaxoSmithKline, Glenmark, MediciNova, Merck Serono, Novartis, Novartis Research Foundation, Roche, Roche Research Foundation, Sanofi-Aventis, Santhera, Swiss MS Society, Swiss National Research Foundation, Teva Neuroscience, UCB, and Wyeth.

Robert Hyde is an employee of Biogen.

Shibeshih Belachew is an employee of Biogen.

Freek Verheul is an advisory board member for Teva Biogen Merck Serono and Novartis.

Francois Grand-Maison received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.
Guillermo Izquierdo received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

Helmut Butzkueven received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital.

References:


Comparative analysis of MS disability regression independent of relapse recovery in natalizumab-treated patients using multinomial propensity score matching

Title: Comparative analysis of MS disability regression independent of relapse recovery in natalizumab-treated patients (paper-format report)

AUTHORS: Tim Spelman,\textsuperscript{1} Fabio Pellegrini,\textsuperscript{2} Annie Zhang,\textsuperscript{2} Maria Trojano,\textsuperscript{3} Heinz Wiendl,\textsuperscript{4} Ludwig Kappos,\textsuperscript{5} Robert Hyde,\textsuperscript{2} Shibeshih Belachew,\textsuperscript{2} Freek Verheul,\textsuperscript{6} Francois Grand-Maison,\textsuperscript{7} Guillermo Izquierdo,\textsuperscript{8} Helmut Butzkueven,\textsuperscript{1} on behalf of the MSBase Substudy Investigators and the TOP Investigators

\textsuperscript{1}Department of Medicine and Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia
\textsuperscript{2}Biogen Inc., Weston, MA
\textsuperscript{3}Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy
\textsuperscript{4}Department of Neurology, University of Münster, Münster, Germany
\textsuperscript{5}Departments of Neurology and Biomedicine, University Hospital Basel, Basel, Switzerland
\textsuperscript{7}Groene Hart Ziekenhuis, Gouda, The Netherlands
\textsuperscript{6}Neuro Rive-Sud, Hôpital Charles LeMoyne, Quebec, Canada
\textsuperscript{8}Hospital Universitario Virgen Macarena, Sevilla, Spain

Corresponding Author: Tim Spelman

Postal Address:
Melbourne Brain Centre, Royal Melbourne Hospital,
Grattan St, Parkville, Victoria, Australia 3050
Ph: +61 3 9342 8070
Timothy Denis Spelman (58172)

Fax: +61 3 9342 8070

Email: tim@burnet.edu.au

**Keywords:** natalizumab, interferon-beta, glatiramer acetate, fingolimod, teriflunomide, multinomial propensity-matching, treatment switching, relapse, disability progression
Background

Many comparative head-to-head analyses of treatment efficacy between competing DMD product options focus on disability progression as a primary end-point. Traditional binary progression metrics, including those featured in pivotal MS clinical trials often do not take account of improvement or regression in disability secondary to treatment which may be secondary to recovery following a relapse event or even improvement independent or unrelated to a clinical relapse.

The objective of this analysis is to compare relapse and progression outcomes across a range of treatment switch scenarios. The analysis describes a novel application of multinomial propensity score matching across three separate switch treatment groups.

Methods

Data sources

Standard Protocol Approvals, Registrations, and Patient Consents Patients treated with BRACE or natalizumab were extracted from the MSBase and Tysabri Observational Programme registries respectively. Details of these registries have been previously published.\textsuperscript{16,24} In both registries, relapse is characterised according to McDonald’s criteria,\textsuperscript{25} whilst interactive Neurostatus training is used for consistent assessment of EDSS. Project approval from an ethical standards committee on human experimentation (institutional or regional) for any experiments using human subjects was obtained by each clinical centre contributing data to either MSBase or TOP registries. Written informed consent was obtained from all patients contributing data to either registry in accordance to the local regulations and laws applicable at each clinic.

Study design

Treatment switch groups
Three treatment switch patient groups were specified for the comparative three-way analysis: 1) relapsing remitting MS (RRMS) patients switching from Interferonβ (IFNβ), glatiramer acetate (GA), teriflunomide (T) or dimethyl fumarate (DMF) (referred overall from here on as IFNβ/GATD for convenience) to natalizumab; 2) RRMS patients switching between IFNβ/GATD preparations and 3) RRMS patients switching from IFNβ/GATD to fingolimod.

**Inclusion and exclusion criteria**

For all three switch treatment groups, the following conditions were required to be satisfied at the time of commencing the switch treatment (denoted as baseline): 1) EDSS ≥2.0 at baseline; 2) no relapse events of worsening of EDSS within a 12 week interval prior to baseline; 3) at least 1 relapse within the 24 months pre-switch baseline and 4) recording a minimum follow-up of 6 months of continuous therapy on the switch product. Patients previously exposed to alemtuzumab were excluded from the analysis.

**Study end-points**

For proof-of-concept, annualised relapse rate (ARR), time to first on-treatment relapse and time to first three month confirmed disability progression were analysed as end-points. Confirmed disability progression events were defined as a minimum 3-month confirmed increases of ≥0.5, ≥1.0 and ≥1.5 points for patients with baseline EDSS >5.5, between 1.0 and 5.5 and those with baseline EDSS of 0 respectively. EDSS scores recorded within 30 days post-relapse were excluded. The comparison of disability progression by treatment arm was limited to patients contributing at least 3 EDSS assessment points, as the minimum required to first observed and then confirm a progression event.

Once the feasibility of the multinomial PSM approach was established, the analysis then considered the following primary study outcomes:
1. Propensity score matched comparison between the 3 switch populations of time to 3-month confirmed EDSS regression.

2. Propensity score matched comparison between the 3 switch populations of time to 6-month confirmed EDSS regression.

3. Propensity score matched comparison between the 3 switch populations of time to 12-month confirmed EDSS regression.

4. Sensitivity analysis using pair-wise censoring at relapse occurrence for all aforementioned endpoints.

5. Subgroup analysis will be performed by discreet categories of baseline (Time 0) EDSS and disease duration.

Statistical analyses

Categorical variables were summarized using frequency and percentage and continuous factors summarised using median and inter-quartile range (IQR) or mean and standard deviation (SD) as appropriate. Eligible patients satisfying the inclusion criteria for each of the three treatment switch groups were propensity score matched using a multinomial 1:1:1 design. The propensity score was derived as a function of the following baseline characteristics as independent, explanatory covariates: sex, age at switch, baseline EDSS, disease duration, number of separate pre-baseline disease-modifying drug (DMD) initiations, the proportion of disease duration spent on DMD therapy, the number of pre-baseline DMD commencements as a proportion of disease duration and pre-baseline relapse activity. The three treatment switch groups were propensity score matched generalised multinomial procedure previously described by Rassen et al.1 Under the procedure propensity score for each member of the three treatment groups is estimated via multinomial logistic regression.2,3 “Within-trio” optimised matching, again without replacement as utilised in the binomial matching described above, is then applied to identify triplets of patients, one from each of the three treatment arms, where the distance in propensity score across the triplet is minimised.
Annualised relapse rate (ARR) between the three matched treatment switch groups were compared using a Friedman test, whilst pair-wise comparisons were conducted using a signed-rank test with Bonferroni adjustment for multiple comparisons. Time to first relapse, first confirmed disability progression and three/six/twelve-month confirmed EDSS regression post-switching were analysed using a Cox Marginal Model. Simultaneous censoring of the matched triplet was applied to adjust for differences in follow-up duration between the switch groups. This involves censoring all three members of the matched triplet at the earliest recorded censor point of any one of the three triplet members. Non-simultaneous censoring of the matched triplet was modelled as a sensitivity analysis. All modelling based on the 489 matched triplets sample were covariate adjusted for pre-baseline relapse activity secondary to significant residual imbalance in this variable between switch treatment arms post-matching. For all models, hazard proportionality was assessed via analysis of scaled Schoenfeld residuals. As a sensitivity analysis, the relapse, progression and regression models were rerun over a smaller matched sample of 382 matched triplets where all observed baseline covariates were well balanced at baseline. A Rosenbaum sensitivity analysis was conducted to test each of the models for the influence of unmeasured confounding. All analyses were conducted in Stata version 13 (StataCorp, College Station, TX) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

The application of inclusion criteria identified a total of 684, 1079 and 489 eligible individual patients from the natalizumab, IFNβ/GATD and fingolimod switch treatment groups respectively. Significant differences between the three switch groups were observed for all baseline characteristics used in the propensity match with the exception of sex (p=0.0970) (Table 1).
Table 1: Comparison of baseline characteristics across the unmatched treatment switch groups

<table>
<thead>
<tr>
<th>Baseline factor</th>
<th>Group 1 (n=684)</th>
<th>Group 2 (n=1079)</th>
<th>Group 3 (n=489)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex - n (%)</td>
<td>484 (70.8)</td>
<td>811 (75.2)</td>
<td>351 (71.8)</td>
<td>0.0970</td>
</tr>
<tr>
<td>Age (years) - median (IQR)</td>
<td>38.50 (31.85, 44.60)</td>
<td>39.62 (33.15, 46.44)</td>
<td>39.12 (32.90, 45.88)</td>
<td>0.0154</td>
</tr>
<tr>
<td>EDSS - median (IQR)</td>
<td>4 (2.5, 5)</td>
<td>3 (2, 4)</td>
<td>2.5 (2, 4)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Disease duration (years) - median (IQR)</td>
<td>9.12 (4.73, 14.23)</td>
<td>8.15 (4.16, 13.42)</td>
<td>9.27 (4.75, 15.35)</td>
<td>0.0035</td>
</tr>
<tr>
<td>Proportion of disease duration on treatment - mean (SD)</td>
<td>0.53 (0.26)</td>
<td>0.44 (0.27)</td>
<td>0.58 (0.26)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Number of DMD treatment starts - mean (SD)</td>
<td>1.90 (1.20)</td>
<td>1.48 (1.19)</td>
<td>3.06 (5.79)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Number of DMD treatment starts / disease duration - mean (SD)</td>
<td>0.29 (0.27)</td>
<td>0.29 (0.36)</td>
<td>0.39 (0.58)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean(SD)</td>
<td>1.33 (0.82)</td>
<td>0.90 (0.83)</td>
<td>1.04 (0.88)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean(SD)</td>
<td>2.28 (1.24)</td>
<td>1.76 (1.06)</td>
<td>1.82 (1.07)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>0.93 (0.79)</td>
<td>0.54 (0.69)</td>
<td>0.75 (0.87)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean(SD)</td>
<td>1.55 (1.17)</td>
<td>0.96 (0.95)</td>
<td>1.13 (1.16)</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

All four-hundred and eighty-nine patients in the smallest pre-matching treatment switch group (the fingolimod switchers) were matched on a 1:1:1 basis to a comparable pair of natalizumab and IFNβ/GATD switchers. In contrast to the unmatched sample baseline factors were well balanced between the three matched treatment arms for all characteristics with the exception of the number of relapses recorded in the 12 and 24 months prior to the switch baseline. For this reason the Cox marginal modelling of time to first relapse and disability progression were covariate adjusted for total relapse count in the 24 months pre-baseline (Table 2).
Table 2: Comparison of baseline characteristics across the propensity matched treatment switch groups

<table>
<thead>
<tr>
<th>Baseline factor</th>
<th>Group 1 (n=489)</th>
<th>Group 2 (n=489)</th>
<th>Group 3 (n=489)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex - n (%)</td>
<td>351 (71.8)</td>
<td>357 (73.0)</td>
<td>351 (71.8)</td>
<td>0.8850</td>
</tr>
<tr>
<td>Age (years) - median (IQR)</td>
<td>39.59 (32.69)</td>
<td>39.00 (32.43, 44.80)</td>
<td>39.12 (32.90, 45.88)</td>
<td>0.6626</td>
</tr>
<tr>
<td>EDSS - median (IQR)</td>
<td>3.5 (2.5, 4)</td>
<td>3.5 (2.5, 4)</td>
<td>3.5 (2.5, 4)</td>
<td>0.3744</td>
</tr>
<tr>
<td>Disease duration (years) - median (IQR)</td>
<td>9.69 (4.86, 14.69)</td>
<td>8.87 (4.86, 13.87)</td>
<td>9.27 (4.75, 15.34)</td>
<td>0.5863</td>
</tr>
<tr>
<td>Proportion of disease duration on treatment - mean (SD)</td>
<td>0.56 (0.26)</td>
<td>0.55 (0.27)</td>
<td>0.58 (0.26)</td>
<td>0.1088</td>
</tr>
<tr>
<td>Number of DMD treatment starts - mean (SD)</td>
<td>1.93 (1.21)</td>
<td>1.71 (1.15)</td>
<td>2.05 (1.24)</td>
<td>0.1937</td>
</tr>
<tr>
<td>Number of DMD treatment starts / disease duration - mean (SD)</td>
<td>0.29 (0.28)</td>
<td>0.27 (0.28)</td>
<td>0.29 (0.28)</td>
<td>0.2258</td>
</tr>
<tr>
<td>Total relapse onsets last 12 months – mean (SD)</td>
<td>1.21 (0.74)</td>
<td>1.15 (0.89)</td>
<td>1.04 (0.88)</td>
<td>0.0013</td>
</tr>
<tr>
<td>Total relapse onsets last 24 months – mean (SD)</td>
<td>2.04 (1.05)</td>
<td>1.95 (1.23)</td>
<td>1.82 (1.07)</td>
<td>0.0003</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean (SD)</td>
<td>0.85 (0.72)</td>
<td>0.78 (0.77)</td>
<td>0.75 (0.87)</td>
<td>0.1087</td>
</tr>
<tr>
<td>Total steroid treated relapse onsets last 12 months – mean (SD)</td>
<td>1.35 (1.03)</td>
<td>1.25 (1.04)</td>
<td>1.13 (1.03)</td>
<td>0.0693</td>
</tr>
</tbody>
</table>

Relapse

Annualised relapse rate (ARR) was significantly higher in the IFNβ/GATD to IFNβ/GATD switch group relative to either the natalizumab or fingolimod switch groups (p<0.0001) (Table 3).

Table 3: Comparison of ARR across the three matched treatment switch groups

<table>
<thead>
<tr>
<th>Switch group category</th>
<th>Mean (SD) ARR</th>
<th>Median (IQR) ARR</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRACED TD to BRACETD</td>
<td>0.60 (0.80)</td>
<td>0.29 (0.00, 0.92)</td>
<td></td>
</tr>
<tr>
<td>BRACED TD to NAT</td>
<td>0.28 (0.53)</td>
<td>0.00 (0.00, 0.40)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BRACED TD to FTY</td>
<td>0.28 (0.56)</td>
<td>0.00 (0.00, 0.37)</td>
<td></td>
</tr>
</tbody>
</table>

* BRACED TD to BRACETD vs BRACETD to NAT: p<0.0001 (Bonferroni adjusted)
* BRACED TD to BRACETD vs BRACETD to FTY: p<0.0001 (Bonferroni adjusted)
* BRACETD to NAT vs BRACETD to FTY: p=0.3473 (Bonferroni adjusted)
Natalizumab switchers were associated with a 47% reduction in the rate of first relapse post-switch relative to patients switching between IFNβ/GATD (adjusted HR: 0.53; 95% CI 0.40, 0.69). Similarly fingolimod switchers were associated with a 51% reduction in the rate of first relapse post-switch relative to patients switching between IFNβ/GATD (adjusted HR: 0.49; 95% CI 0.37, 0.64), adjusting for the count of relapses in the 24 months prior to baseline (Figure 1).

Figure 1: Time to first post-switch relapse by switch treatment group

Disability progression
Natalizumab switchers were associated with a 39% decrease in the rate of first confirmed disability progression relative to IFNβ/GATD switchers (aHR: 0.61; 95% CI 0.39, 0.94). There was no difference between fingolimod switches and patients switching between IFNβ/GATD products (aHR 0.77; 95% CI 0.51, 1.16) (Figure 2).
Time to three-month confirmed EDSS regression

Both natalizumab and fingolimod switchers were associated with an increased rate of three-month confirmed EDSS regression relative to inter-BRACETD switchers. Patients switching to natalizumab were associated with 1.84 times the rate of three-month confirmed EDSS regression relative to BRACETD switchers (HR 1.84; 95% CI 1.35, 2.52) whilst fingolimod switchers were associated with 1.62 times the rate (HR 1.62; 95% CI 1.17, 2.26). Similar effects were observed under simultaneous censoring of the matched triplet with natalizumab and fingolimod switchers correlating with 1.92 (HR 1.92; 95% CI 1.28, 2.87) and 1.86 (HR 1.86; 95% CI 1.23, 2.79) the rate of three-month confirmed regression respectively, again relative to inter-BRACETD switchers.
**Time to six-month confirmed EDSS regression**

Consistent with the three-month confirmed regression modelling, natalizumab and fingolimod switchers were associated with an increased rate of six-month confirmed EDSS regression relative to inter-BRACETD switchers, under both non-simultaneous and simultaneous censoring. Patients switching to natalizumab were associated with 1.90 times the rate of six-month confirmed EDSS regression relative to BRACETD switchers (HR 1.90; 95% CI 1.26, 2.54) whilst fingolimod switchers were associated with 1.79 times the rate (HR 1.62; 95% CI 1.26, 2.54). Marginally larger effects were again observed under simultaneous censoring of the matched triplet with natalizumab and fingolimod switchers correlating with 2.10 (HR 2.10; 95% CI 1.35, 3.27) and 2.00 (HR 2.00; 95% CI 1.28, 3.12) the rate of six-month confirmed regression respectively, again relative to inter-BRACETD switchers.

**Time to twelve-month confirmed EDSS regression**

Similarly consistent with both the three- and six-month confirmed regression modelling, natalizumab and fingolimod switchers were associated with an increased rate of twelve-month confirmed EDSS regression relative to inter-BRACETD switchers, under both non-simultaneous and simultaneous censoring. Patients switching to natalizumab were associated with 2.05 times the rate of twelve-month confirmed EDSS regression relative to BRACETD switchers (HR 2.05; 95% CI 1.42, 2.95) whilst fingolimod switchers were associated with 1.88 times the rate (HR 1.88; 95% CI 1.29, 2.74). Marginally larger effects were again observed under simultaneous censoring of the matched triplet with natalizumab and fingolimod switchers correlating with 2.38 (HR 2.38; 95% CI 1.47, 3.82) and 2.29 (HR 2.29; 95% CI 1.42, 3.70) the rate of twelve-month confirmed regression respectively, again relative to inter-BRACETD switchers.
Subgroup analysis

Median baseline EDSS

The three, six and twelve-month confirmed EDSS regression modelling described above was repeated for sub-groups of the matched sample based on median baseline EDSS (i.e. EDSS at the time of switching). The sample was split into baseline EDSS≤3 and >3 sub-groups.

Subgroup = Baseline EDSS ≤3

Within the baseline EDSS≤3 sub-group, natalizumab switchers were associated with an increased rate of twelve-month confirmed regression (HR 1.71; 95% CI 1.01, 2.88, non-simultaneous censoring) relative to inter-BRACED switchers. There was no difference in three- or six-month confirmed EDSS regression in this sub-group under non-simultaneous censoring, although natalizumab switching was associated with 1.92 times the rate of three-month confirmed EDSS regression using simultaneous censoring (HR 1.92; 95% CI 1.03, 3.60; reference=BRACETD). By comparison fingolimod switchers were associated with an increase in the rate of three- (HR 1.82; 95% CI 1.20, 2.75), six- (HR 1.80; 95% CI 1.15, 2.82) and twelve-month (HR 2.02; 95% CI 1.22, 3.34) confirmed EDSS regression, relative to the inter-BRACETD arm. This pattern was repeated under simultaneous censoring with fingolimod switchers associated with increased rates of three-month (HR 2.56; 95% CI 1.43, 4.58), six-month (HR 2.27; 95% CI 1.23, 4.19) and twelve-month (HR 2.74; 95% CI 1.39, 5.39) confirmed EDSS regression relative to BRACETD switchers.

Subgroup = Baseline EDSS >3

In contrast, the fingolimod signal observed within the ≤3 sub-group disappears in the analysis of the baseline EDSS>3 group. No differences were observed in the rate of three-, six- or twelve-month confirmed EDSS regression in the fingolimod switchers relative to inter-BRACETD switchers, under either non-simultaneous or simultaneous censoring. A far stronger natalizumab signal was observed in this sub-group with natalizumab switchers consistently associated with increased rates of three-,
Timothy Denis Spelman (58172)

six- and twelve-month confirmed EDSS regression under both non-simultaneous and simultaneous censoring. Switching to natalizumab was associated with $2.20$ (HR $2.20; 95\% \text{ CI } 1.39, 3.48$), $2.41$ (HR $2.41; 95\% \text{ CI } 1.42, 4.08$) and $2.42$ (HR $2.42; 95\% \text{ CI } 1.40, 4.16$) times the rate of three-, six- and twelve-month confirmed regression relative to inter-BRACETD switchers under non-simultaneous censoring. Using simultaneous censoring the equivalent effects were HR $1.85$ (95% CI 1.10, 3.12), HR $2.24$ (95% CI 1.23, 4.08) and HR $2.58$ (95% CI 1.37, 4.85). Refer to tab labelled “Median EDSS_489” in the master results spreadsheet for full details.

**Median disease duration**

A similar sub-group analysis was run using median disease duration at baseline as the sub-group filter. The sample was split into baseline disease duration <10 years and ≥10 years subgroups.

*Subgroup: Median baseline disease duration < 10 years*

Both natalizumab and fingolimod switchers were associated with increased rates of three-, six- and twelve-month confirmed EDSS regression relative to the inter-BRACETD switchers. Natalizumab was associated with $1.72$ (HR $1.72; 95\% \text{ CI } 1.16, 2.57$), $1.78$ (HR $1.78; 95\% \text{ CI } 1.15, 2.76$) and $1.90$ (HR $1.90; 95\% \text{ CI } 1.19, 3.03$) times the rate of three-, six- and twelve-month confirmed EDSS regression respectively under non-simultaneous censoring and $1.91$ (HR $1.91; 95\% \text{ CI } 1.14, 3.21$), $2.27$ (HR $2.27; 95\% \text{ CI } 1.30, 3.96$) and $2.43$ (HR $2.43; 95\% \text{ CI } 1.33, 4.46$) times the rate of three-, six- and twelve-month confirmed EDSS regression respectively using simultaneous censoring.

*Subgroup: Median baseline disease duration ≥ 10 years*

Natalizumab was also observed to correlate with increased rates of regression in the ≥10 years disease duration sub-group. Switching to natalizumab was associated with $2.07$ (HR $2.07; 95\% \text{ CI } 1.25, 3.42$), $2.12$ (HR $2.12; 95\% \text{ CI } 1.22, 3.68$) and $2.38$ (HR $2.38; 95\% \text{ CI } 1.28, 4.44$) times the rate of three-, six- and twelve-month confirmed EDSS regression respectively relative to inter-BRACETD.
switchers under non-simultaneous censoring and 1.94 (HR 1.94; 95% CI 1.02, 3.69) and 2.33 (HR 2.33; 95% CI 1.07, 5.06) times the rate of three- and twelve-month regression under simultaneous censoring. In contrast to the results observed in the <10 years duration sub-group, fingolimod was not associated with a change in any of three-, six- and twelve-month regression rate under with non-simultaneous or simultaneous censoring. Refer to tab labelled “Median DISDURN_489” in the master results spreadsheet for full details.

**Sensitivity analysis**

The relapse, progression and regression end-points were re-examined on a smaller sample of 382 matched triplets demonstrating superior baseline balance in prognostic correlates at baseline. Consistent with the primary analysis, both natalizumab (HR 0.51; 95% CI 0.41, 0.63) and fingolimod (HR 0.53; 95% CI 0.43, 0.66) switchers were again associated with reduced rates of first relapse, relative to the inter-BRACETD switchers. No difference in time to first three-month confirmed disability progression by switch treatment arm was observed. Also consistent with the primary analysis natalizumab was again correlate with 1.92 (HR 1.92; 95% CI 1.37, 2.70), 1.90 (HR 1.90; 95% CI 1.33, 2.71) and 1.95 (HR 1.95; 95% CI 1.33, 2.85) times the rate of three-, six- and twelve-month confirmed EDSS regression respectively, relative to inter-BRACETD switchers. Similarly, fingolimod switchers were associated with 1.72 (HR 1.72; 95% CI 1.20, 3.46), 1.85 (HR 1.85; 95% CI 1.25, 2.74) and 1.89 (HR 1.89; 95% CI 1.25, 2.86) times the the rate of three-, six- and twelve-month confirmed EDSS regression respectively, again relative to inter-BRACETD switchers. Refer to tab labelled “PAIRS 382 RESULTS” in the master results spreadsheet for full details.
Author contributions:

Tim Spelman conceptualized and designed the study, programmed and conducted the statistical analysis, interpreted the analysis and drafted and revised the report

Fabio Pelligrini interpreted the analysis and revised the report

Annie Zhang interpreted the analysis and revised the report

Maria Trojano interpreted the analysis and revised the report

Heinz Wiendl interpreted the analysis and revised the report

Ludwig Kappos interpreted the analysis and revised the report

Robert Hyde interpreted the analysis and revised the report

Shibesih Belachew interpreted the analysis and revised the report

Freek Verheul interpreted the analysis and revised the report

Francois Grand-Maison interpreted the analysis and revised the report

Guillermo Izquierdo interpreted the analysis and revised the report

Helmut Butzkueven conceptualized and designed the study and revised the report

Author disclosures:

Tim Spelman received compensation for serving on scientific advisory boards, honoraria for consultancy and funding for travel from Biogen Inc; speaker honoraria from Novartis.

Fabio Pelligrini is an employee of Biogen.

Annie Zhang is an employee of Biogen.

Maria Trojano received honoraria for consultancy and/or speaking from Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis, and Roche; research grants from Biogen Idec, Merck Serono, Novartis, and Teva.

Heinz Wiendl received compensation for serving on scientific advisory boards for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi; speaker honoraria and travel support from Bayer Schering AG, Bayer Vital GmbH, Biogen Idec, CSL Behring, Fresenius Medical Care,
Timothy Denis Spelman (58172)

Genzyme, GlaxoSmithKline, GW, Merck Serono, Novartis, and Sanofi; compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi; research support from Bayer Vital GmbH, Biogen Idec, Merck Serono, Novartis, Sanofi Germany, and Sanofi US.

**Ludwig Kappos** received research support from Acorda, Actelion, Allozyne, BaroFold, Bayer HealthCare, Bayer Schering, Bayhill Therapeutics, Biogen Idec, Elan, European Union, Genmab, Gianni Rubatto Foundation, GlaxoSmithKline, Glenmark, MediciNova, Merck Serono, Novartis, Novartis Research Foundation, Roche, Roche Research Foundation, Sanofi-Aventis, Santhera, Swiss MS Society, Swiss National Research Foundation, Teva Neuroscience, UCB, and Wyeth.

**Robert Hyde** is an employee of Biogen.

**Shibeshih Belachew** is an employee of Biogen.

**Freek Verheul** is an advisory board member for Teva Biogen Merck Serono and Novartis.

**Francois Grand-Maison** received an honorarium for organizing a CME event for Biogen Idec in 2013 and received consultation fees from Biogen Idec as well as from Novartis and Genzyme in 2013 and 2014.

**Guillermo Izquierdo** received consulting fees from Bayer Schering, Biogen Idec, Merck Serono, Novartis, Sanofi, and Teva.

**Helmut Butzkueven** received compensation for serving on scientific advisory boards and as a consultant for Biogen Idec and Novartis; speaker honoraria from Biogen Idec Australia, Merck Serono Australia, and Novartis Australia; travel support from Biogen Idec Australia and Merck Serono Australia; research support from CASS Foundation (Australia), Merck Serono Australia, the Royal Melbourne Hospital.
H. FINAL DISCUSSION

H.1 Summary of findings

This thesis describes the novel derivation and application of a series of statistical techniques for analysing observational outcome data sourced from a large data registry of multiple sclerosis patients. Section D describes a novel application and evaluation of trigonometric regression modelling to 32,762 relapse events sourced from 46 clinical centres across 20 countries spanning both northern and southern hemispheres to show for the first time that the lag between seasonal UVR winter trough and subsequent peak in MS relapse frequency varies with latitude away from the equator. The modelling presented also confirms existing meta-analyses describing a northern hemispheric seasonality in the timing of relapse onset. It further extends this observation, again for the first time, to the southern hemisphere – a region largely ignored by the precedent literature. A regression model was used to model relapse onset and ultra-violet radiation (UVR) seasonality specified as a function of a series of trigonometric functions to capture seasonal trends. Linear regression was used to investigate associations of latitude and lag between UVR trough and subsequent relapse peak. Relapse onset followed an annual cyclical sinusoidal pattern with peaks in early spring and troughs in autumn in both hemispheres. Every 10 degrees of latitude away from the equator was associated with a mean decrease in ultra-violet radiation trough to subsequent relapse peak lag of 28.5 days (95% CI 3.29, 53.71, p=0.028).

Chapter E presents a prospective analysis of correlates and predictors of conversion to clinically definite MS (CDMS) in the, to date, largest studied seen from onset clinically isolated syndrome (CIS) international cohort available. A total of 3296 patients from 50 clinics in 22 countries were followed up for a median (IQR) of 1.92 years (0.90, 3.71). Of these, a total of 1953 (59.3%) patients recorded a second attack, marking conversion to CDMS. Cox proportional hazards regression was used to identify age at CIS, higher EDSS, brainstem or supratentorial first symptom location, the presence of oligoclonal bands, 1+ T1 gadolinium enhancing lesions, 3+ periventricular lesions, 1+ infratentorial
and 1+ juxtacortical lesion on baseline MRI were as independent correlates of subsequent conversion risk. These adjusted model results were then used to derive a series of internally-validated prognostic conversion risk nomogram tools, designed to permit personalised conversion risk assessment at the time of CIS in clinical practice.

Section F flags the first in a series of treatment outcome analyses, in this case the important clinical question of which demographic, disease activity or examination characteristics predict premature treatment discontinuation. Cox time-to-event modelling was employed to study n=2003 first-line disease-modifying drug initiations in a prospective, wholly seen-from-onset cohort. Female sex, higher baseline EDSS, higher on-treatment annualised relapse rate (ARR) were all independently correlate with an increased rate of first-line DMD product cessation. Marked country effects were further observed with patients managed in Australian clinics correlating with higher discontinuation rates compared to patients managed in European or North American clinic settings. Differential discontinuation rates by DMD product were further observed with patients commenced on first line intra-muscular (IM) IFNβ-1a discontinuing at a significantly higher rate than any of the other three platform injectable medications (sub-cutaneous IFNβ-1a, IFNβ-1b or glatiramer acetate). Conversely, on-treatment EDSS change, age at either disease onset or first DMT initiation and baseline cerebral MRI parameters were not association with discontinuation.

Section G presents a series of head-to-head treatment product and decision point comparisons using a range of propensity-score based matching and adjustment procedures. Section G.1 presents a matched comparison of two common treatment switch scenarios: switching between IFNβ and GA, compared with switching to natalizumab in a cohort of patients with active disease on treatment. Switching to natalizumab was associated with a significant reduction in ARR in year one by 65-75%, the risk of first relapse by 53-82% (mean follow-up 1.7-2.2 years) and treatment discontinuation events by 48-65% (all \( p \leq 0.001 \)). Switching to natalizumab reduced the risk of confirmed disability
progression by 26% ($p=0.036$) and decreased the total disability burden by 1.54 EDSS-years [$p<0.0001$]) over the first 24 months post-switch.

Moving from a switch scenario to first-line setting, natalizumab as a first DMD commencement was then compared against propensity matched first-line IFNβ/GA to study a similar suite of treatment end-points (section G.2). First-line natalizumab was associated with a 68% relative reduction in ARR from a mean (SD) of 0.63 (0.92) on IFNβ/GA to 0.20 (0.63) ($p(\text{signed-rank})<0.0001$), a 64% reduction in the rate of first relapse (hazard ratio [HR]= 0.36; 95% confidence interval [CI]=0.28-0.47; $p<0.001$), and a 27% reduction in the rate of discontinuation (HR=0.73, 95%CI=0.58-0.93; $p=0.01$), compared with first-line IFNβ/GA therapy. Confirmed disability progression and area under the EDSS-time curve analyses were not statistically significant. Similar relapse and treatment persistence results were observed in each of the higher disease activity subgroups studied.

Section G.3 expands the analyses beyond head-to-head efficacy comparisons of established platform injectable and natalizumab infusion treatments to consider new era oral immunomodulatory agents. Specifically this section describes a model for early relapse or rebound disease in n=396 previously stable patients switching from injectable therapy to oral agents, relative to a 1:1 propensity matched cohort of similarly stabilised patients who remain on injectable therapy. No difference in the proportion of patients recording at least 1 relapse in the first 1-6 months by treatment arm (7.3% of switchers; 6.6% stayers, $p=0.675$). The mean Annualised Relapse Rate ($p= 0.493$) and the rate of first 6-month relapse by treatment arm (HR 1.22, 95% CI 0.70, 2.11) were also comparable. There was no difference in the rate of disability progression by treatment arm (HR 1.43, 95% CI 0.63, 3.26).

Section G.4 then directly compares a series of treatment efficacy and persistence end-points in first line fingolimod, a new era oral DMD, with the first-line injectables (IFNβ/GA) in a cohort of patients with active disease using a 2:1 variant of the propensity matching described in the sections
summarised above. Mean (SD) on-treatment ARR on n=229 first-line fingolimod treated patients was decreased by 43% to 0.28 (0.76) relapses per year from 0.49 (0.84) on n=458 matched IFNβ/GA treatments (p(signed-rank)<0.0001). Fingolimod was associated with a 40% reduction in the rate of first on-treatment relapse compared with first-line IFNβ/GA (HR 0.60, 95% CI 0.39, 0.94). First line fingolimod was associated with a 49% reduction in treatment discontinuation compared with IFNβ/GA (HR 0.51, 95% CI 0.37, 0.70). There was no difference in the rate of three-month confirmed disability progression (HR 1.95; 95% CI 0.85, 4.50; reference = IFNβ/GA).

Section G.5 employs a similar matching and modelling approach to combine and compare data from two distinct MS registry based sources – MSCOMET (a substudy of the MSBase registry) and the Tysabri Observational Programme (TOP) safety and efficacy registry. Patients with active disease switching to natalizumab from IFNβ/GA were associated with a 64% decrease in the rate of on-treatment relapse (HR 0.36, 95% CI 0.30, 0.44), a 44% decrease in the rate of treatment discontinuation (HR 0.56, 95% CI 0.50, 0.63) and a 39% reduction in the rate of confirmed disability progression (HR 0.61, 95 CI: 0.36, 0.51) relative to 2:1 propensity matched cohort of active disease patients persisting on IFNβ/GA following at least 1 on-treatment relapse event. The switch group was further associated with a greater mean (SD) decrease in on-treatment EDSS (-0.04 (0.98) vs 0.06 (1.08); p=0.0159).

This series of propensity matched efficacy and persistence comparisons concludes with the presentation of a multinomial PS match across three treatment comparison groups (section G.6). A total of 489 patients switching from IFNβ/GA/Teriflunomide/Dimethyl Fumarate (IFNβ/GATD) to fingolimod were matched a 1:1:1 basis to a comparable pair of natalizumab switchers and patients switching between IFNβ/GATD preparations. Annualised relapse rate (ARR) was significantly higher in the IFNβ/GATD to IFNβ/GATD switch group relative to either the natalizumab or fingolimod switch groups (p<0.0001). Natalizumab switchers were associated with a 47% reduction in the rate of first
relapse post-switch relative to patients switching between IFNβ/GATD (adjusted HR: 0.53; 95% CI 0.40, 0.69). Similarly fingolimod switchers were associated with a 51% reduction in the rate of first relapse post-switch relative to patients switching between IFNβ/GATD (adjusted HR: 0.49; 95% CI 0.37, 0.64). Natalizumab switchers were associated with a 39% decrease in the rate of first confirmed disability progression relative to IFNβ/GATD switchers (aHR: 0.61; 95% CI 0.39, 0.94). There was no difference between fingolimod switches and patients switching between IFNβ/GATD products (aHR 0.77; 95% 0.51, 1.16).

H.2 MS registry data – advantages, opportunities and future directions

What is consistent across all analyses presented in this thesis is the marked statistical power conferred by a large, global, longitudinal registry collecting a standardised minimum dataset for analysing real world outcomes over long time periods. Not only does this permit modelled estimation of treatment effect or MS outcome that better approximate the true population effect size, relative to the smaller datasets that constitute the bulk of the precedent literature, it further allows development, exploration and adaptation of sample-intensive statistical techniques and methodologies necessary for teasing out independent, stand-alone predictors of adverse events or favourable treatment response in a complex, chronic disease such as MS. A consistent finding across all four analytical themes in this thesis is the number of demographic, environmental, clinical, disease activity, treatment and examination factors that correlate with the MS end-points studied, from the seasonal timing of relapse onset, through to drivers and modulators of conversion probability and response to treatment.

Not only do these factors correlate with multiple outcomes, they also interact markedly with one another. This has several important ramifications for how such data is analysed. First is the prerequisite for a sufficient sample size with long follow-up duration. Without this, there is no scope for the application of sophisticated, multivariate and/or multinomial regression-based modelling
techniques. Second is the need for appropriately implemented and evaluated statistical modelling which can simultaneously control for the influence of many competing, and often highly correlated, confounding factors whilst minimising the various biases inherent in observational data, especially in relation to treatment assignments. A clear advantage of a registry data source such as MSBase is its size and internal consistency, allowing the careful application of sophisticated statistical approaches.

This is well illustrated in the example of the trigonometric regression modelling applied to global relapse data as described in section D. Previous observations of a seasonal gradient in relapse timing over multiple geographical locations have relied upon pooling separate studies into a single meta-analytical comparison limited by significant heterogeneity and lack of exchangeability between studies. By comparison, the geometric modelling presented in this thesis, by using data from a single registry source employing a unified data capture and definition protocol, avoids these problems, allow better isolation of genuine seasonal and, in this case, latitudinal correlates of relapse probability. Unlike the majority of previous studies exploring such gradients, the large relapse dataset coupled with the intentionally regression-based modelling approach permitted adjustment for important patient level confounders of these environmental signals including exposure to treatment and disability level. This was a critical step for making inferences around a genuine association between season, latitude and relapse as it permitted quantification of an association entirely independent of both location-level and patient-level effect confounders. The fact that previous studies have either not applied a geometric regression approach or adjusted for patient level confounders may in part explain some of the disagreements in the literature around whether relapse seasonality is real or a statistical artefact. The cure for suspected statistical artefact is good statistical methodology designed to separate and isolate multiple, competing effect modulators. Thus the modelling approach described in section D and further detailed in the methods presented
in section C has potential for wide application for filtering out environmental signals which, specifically, describe predictable, cyclic, periodic fluctuations over time.

This exact same principle was illustrated in the development of prognostic conversion nomograms described in chapter E. The key to the internal validity of the presented nomograms was access to large, wholly prospective, seen from onset MSBASIStudy. The MSBase sub-study system means the internal consistency of the overall registry data is automatically carried over into any sub-studies focusing on particular subsets of patients. This means a large suite of candidate predictors of conversion were available to be tested and simultaneously compared to one another to derive an estimate of the relative magnitude of effect. Through large model covariate adjustment, good isolation of the independent effects of demographic, clinical, disease activity and examination metrics at, typically, time of first presentation on subsequent conversion probability is possible – an event which may occur years down the track. Indeed it was this very combination of dataset size and model selection process that delivered the predictor separation required to then build what are, explicitly, personalised risk assessment nomograms – i.e. tools derived on real world population data for application at the level of the individual data.

Such empirically grounded risk assessment and prediction tools have potentially wide utility in clinical practice, for both clinicians and patients. Nomograms such as those derived and described in section E can feasibly be converted to on-line calculators or incorporated into mobile phone applications to return quick estimates of risk based on the characteristics and circumstances of an individual patient presenting to clinic, and this method of translation of results to clinical practice is planned in 2016. Importantly these risk prediction models and tools can assist both the clinician and patient in negotiating treatment decisions around the optimal timing of commencing therapy – on the basis of the unique set of individual patient factors rather than inferred from broad, aggregate
population data. This may in turn permit better customisation of treatment in a setting where treatment durations in MS are typically long, incompletely efficacious and often poorly tolerated.

Whilst randomised clinical trials remain the gold standard for evidence around comparative treatment efficacy, the various propensity score matching techniques described in this thesis provide a possible solution for returning relatively unbiased estimates of treatment effects, in particular for comparative treatment decision points for which clinical trial data is absent (ie almost all of them). Whilst propensity score matching cannot achieve the same degree of balance as a clinical trial, an appropriately managed and, critically, evaluated match can return reasonably robust approximations of the effect sizes typically observed in the pivotal MS drug trials.

The real value of a propensity based approach, in a lifelong disease such as MS, at least, is its potential application in studying long-term treatment outcomes, particular when applied to a large registry data source. Whilst clinical trials provide strong evidence of comparative efficacy in early treatment, most trials (including various open-label extensions) cover only a comparatively short duration (typically 12 to 24 months) of a MS patient’s total treatment experience. Further, treatment in MS is rarely as neat or as linear as appears in a tightly controlled trial. Many patients will cycle through various sequential treatments secondary to efficacy, tolerability, adverse effects and, being a disease which disproportionately strikes young women, pregnancy. Propensity based methods, when applied to a dataset of sufficient size and follow-up, is potentially a reliable tool for examining the diverse treatment pathways or trajectories. This was in part the rationale for presenting a series of propensity match scenarios in section G, covering a series of treatment start/discontinuation triggers and disease activity patient subsets. An additional advantage of large registry sources is their capacity to quickly accumulate real world outcome data on newer therapies. This is particularly relevant in the current therapeutic climate of recent developments in the trialling
and approval of novel oral immunomodulatory agents for MS, after years where clinicians had access to a small suite of injectable therapies only.

**Future direction**

Whilst the various propensity score matching techniques described herein work to rebalance prognostic correlates of treatment end-points between non-randomly assigned treatment groups, this balance is applied at baseline only – typically the start date of a treatment or an alternate treatment trigger such as discontinuation. Whilst this ensures that competing treatment arms are well balanced at a single point in time, it does not control for systematic differences between treatment arms introduced subsequent to baseline (i.e. whilst on treatment). Whilst these effects can be minimised in a clinical trial setting secondary to tight study protocols and frequent assessment points, they can reintroduce significant imbalance and thus bias into observational cohorts, even after propensity matching at baseline. Recent promising developments in the fusion of propensity score methods with marginal structural models (MSM) may provide an attractive solution to this problem of time-varying confounding. Rather than deriving propensity score as function of baseline confounders only, the MSM approach derives the score longitudinally across the observation period, re-calculating the probability that a patient would theoretically been in receipt of the intervention based on a given suite of prognostic correlates at time $t$. The separate time $t$ scores are then synthesised to generate an overall MSM PS, which can then be used to inversely weight the MSM on the propensity score to control for such time-varying confounding. What is particularly attractive about this approach is its flexibility. By permitting inclusion of lag variables (e.g. lag of treatment histories, lag of confounders) in the derivation of the propensity score it can be used not only to adjust for both baseline and time-varying confounding, it can also be used to study the complex, multi-product treatment pathways and trajectories becoming more common in MS treatment secondary to the expansion of the available therapy base. A dataset such as MSBase
would be well placed to investigate the suitability of MSM based methods, particularly as it continues to accumulate data and follow-up on newer therapies.

H.3 Limitations

As detailed in section D, the seasonal and latitudinal gradients observed and their relationships to ultraviolet radiation are themselves hypothesis generating and do not in themselves constitute evidence for an underlying vitamin D modulated biological mechanism for relapse. Indeed the post-hoc hypothesis, that the observed lag between winter UVR trough and subsequent relapse peak is best explained by an effect of changing vitamin D status and/or other UVR-related direct immune influence with a several-week duration to clinical symptomatology is more appropriately studied in the context of a randomised clinical trial. As described in the discussion of section D, there are currently two trials underway investigating the direct influence of vitamin D supplementation on MS relapse probability in CIS, although these are not due to report until 2018-19.

Whilst the nomograms presented in section E performed well on internal validation (through calculation of concordance indices and derivation of calibration plots), external validation against an independent data source forms a pre-requisite before potentially expanding these tools into clinical practice as risk assessment tools. This would require an independent data source of at least comparable size and power. A potential source for this may present itself in the form of the new “Big MS” collaboration – a recent concern investigating combining multiple large MS registries into a single pooled data sourced. Significant challenges remain to ensure such pooling is not compromised by the same heterogeneity and exchangeability issues faced by many of the meta-analyses covered in the literature review.

As cited several times throughout the thesis, the main limitation of the propensity-matched analyses is that the source data remains observational, subject to the many and varied biases and
inconsistencies characteristic of observational data. As detailed in both the methods and discussion sections of the various matched analyses, a number of checks and balances have been proposed to minimise bias whilst preserving generalisability including the use of conservative censoring regimes (e.g. simultaneous censoring) and targeted sensitivity analyses for estimating the influence of unobserved confounding (e.g. Rosenabum sensitivity analysis). Please refer to section G for a more detailed description and discussion of these techniques including commentary around the likely direction of bias conferred by any residual bias.

H.4 Conclusion

Multiple sclerosis is complex. Whilst a strong evidence base from clinical trials supports early treatment efficacy, these typically capture only a short duration of a patient’s total disease experience. Many studies of observational cohorts investigating independent correlates of complex, multifactorial outcomes such as relapse and disability progression or patient response to treatment are limited by lack of power or an insufficient analytical approach. Large, established observational registry datasets such as MSBase provide a unique opportunity to study complex, long duration epidemiological and treatment outcomes. They further permit development and evaluation of statistical techniques specifically designed to isolate independent effects and correlates of disease. The various techniques described herein have potential application in other complex, chronic diseases.
I. APPENDIX A – PRISMA flowcharts and checklist

I.1 Hypothesis 1 literature review – PRISMA flowchart

**PRISMA 2009 Flow Diagram – Hypothesis 1**: The timing of relapse onset is seasonal and this relationship is latitude-dependent

- **Records identified through database searching** (n = 34)
- **Additional records identified through other sources** (n = 19)
- **Records after duplicates removed** (n = 51)
- **Records screened** (n = 51)
- **Records excluded** (n = 6)
- **Full-text articles assessed for eligibility** (n = 45)
- **Full-text articles excluded** (n = 8)
- **Studies included in qualitative synthesis** (n = 37)

*For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).*
I.2 Hypothesis 2 literature review – PRISMA flowchart

PRISMA 2009 Flow Diagram – Hypothesis 2: Demographic, clinical, examination and disease activity characteristics at the time of clinically isolated syndrome predict future risk of conversion to clinically definite multiple sclerosis.
I.3 Hypothesis 3 literature review – PRISMA flowchart

**PRISMA 2009 Flow Diagram – Hypothesis 3**: Demographic, clinical, examination and disease activity characteristics at treatment initiation and during therapy predict early discontinuation.


For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).
I.4 Hypothesis 4 literature review – PRISMA flowchart

**PRISMA 2009 Flow Diagram – Hypothesis 4**: Propensity-score matching can return unbiased estimates of comparative treatment efficacy across a range of treatment scenarios and products.


For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).
Thesis publications:

In print:


Accepted:

Conference Proceedings:


- PLATFORM PRESENTATION: Spelman, T et al. Independent predictors of time to relapse after CIS in high-risk patients. 2014 Joint ACTRIMS-ECTRIMS meeting, 10-13 September 2014, Boston Massachusetts, USA

- PLATFORM PRESENTATION: Spelman, T et al. Seasonal variation of relapse rate in multiple sclerosis is latitude dependent. Melbourne Health Research Week, October 2014

- POSTER: Spelman, T et al. Seasonal variation of relapse rate in multiple sclerosis is latitude dependent. 2014 Joint ACTRIMS-ECTRIMS meeting, 10-13 September 2014, Boston Massachusetts, USA


Related publications

**Joint first author:**


**Co-author:**


(MS) Immunotherapy Study: A prospective, multicentre study of drug utilisation using the MSBase platform. *PLoS ONE* 01/2013; 8(3):e59694

**Related conference proceedings**

**Joint first author:**

- PLATFORM PRESENTATION: Kister, I., **Spelman, T.**, Alroughani, R., et al. Are stable MS patients who stop their disease-modifying therapy (DMT) at increased risk for relapses and disability progression compared to patients who continue on DMTs? A propensity-score matched analysis of the MSBase registrants. 2015 ECTRIMS meeting, 7-10 October 2015, Bacerlona, Spain.

**Co-author:**


- POSTER: Jokubaitis, V., Spelman, T. et al. Predictors of disability worsening in clinically isolated syndrome. 2014 Joint ACTRIMS-ECTRIMS meeting, 10-13 September 2014, Boston Massachusetts, USA

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
SPELMAN, TIMOTHY

Title:
Multiple Sclerosis MSBase registry: using real-world data to define MS outcomes and optimise treatment strategies

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/108598