BMJ open

UK Multiple Sclerosis Risk-sharing Scheme: developing an improved analysis plan

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004073
Article Type:	Research
Date Submitted by the Author:	19-Sep-2013
Complete List of Authors:	Palace, Jacqueline; John Radcliffe Hospital, Clinical Neurology Bregenzer, Thomas; PAREXEL International, Biostatistics Unit Tremlett, Helen; University Of British Columbia, Medicine (Neurology) Oger, Joel; University Of British Columbia, Division of Neurology Zhu, Feng; University Of British Columbia, Medicine (Neurology) Boggild, Mike; The Townsville Hospital, Neurology Department Duddy, Martin; The Newcastle upon Tyne Hospitals, Neurology Department Dobson, Charles; Department of Health,
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Health economics, Pharmacology and therapeutics, Research methods
Keywords:	Multiple sclerosis < NEUROLOGY, quality of life, risk sharing scheme, interferon-beta, glatiramer acetate, Markov model

UK Multiple Sclerosis Risk-sharing Scheme: developing an improved analysis plan

The lead author* affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests:

Jacqueline Palace serves on the scientific advisory board for Charcot Foundation, and has performed advisory work for Biogen Idec, Merck Serono Ltd, Bayer Schering Pharma, Novartis Pharmaceuticals UK Ltd, Teva Pharmaceutical Industries Ltd, Gilenya, Ono Pharmaceutical Co Ltd, Primary i-research, Chugai Pharma Europe and CI Consulting. She receives research support from the MS Society, QIDIS, Merck Serono Ltd, Novartis Pharmaceuticals and Bayer Schering Pharma, plus conference expenses from Novartis and Merck Serono Ltd.

Thomas Bregenzer – no financial interests to declare

Helen Tremlett is funded by the Multiple Sclerosis Society of Canada (Don Paty Career Development Award), Michael Smith Foundation for Health Research and is the Canada Research Chair for Neuroepidemiology and Multiple Sclerosis. She has received: research support from the US National Multiple Sclerosis Society, CIHR, and UK MS Trust; speaker honoraria and/or travel expenses to attend conferences from the Consortium of MS Centres, US National MS Society, the University of British Columbia Multiple Sclerosis Research Program, Bayer Pharmaceutical (speaker, 2010, honoraria declined), Teva Pharmaceuticals (speaker 2011), ECTRIMS (2011, 2012), UK MS Trust (2011), and the Chesapeake Health Education Program, US Veterans Affairs (2012, honorarium

declined), Novartis Canada (2012) and the US National MS Society (2012, honorarium declined). Unless otherwise stated, all speaker honoraria are either donated to an MS charity or to an unrestricted grant for use by her research group.

Joel Oger is funded by the University of British Columbia and has received grants and contracts from Biogen-Idec, Bayer, EMD Serono, Merk Serono. He has received speaker fees, Advisory Board membership and travel grants from Bayer, Novartis, Serono and the Ministry of Health of British Columbia. His lab is partially funded by the Christopher Foundation .

Feng Zhu – no financial interests to declare

Mike Boggild sits on advisory boards for Bio CSL, genzyme & Biogen Idec. Recieved sponsorship to attend international meetings from Novartis & BioCSL. Department has received funding to develop services from Biogen Idec, Genzyme and Novartis.

Martin Duddy over the past 5 years MD has received speaker honoraria, consulting fees and travel grants from, Bayer, BiogenIdec, Novartis, Merck-Serono and Teva

Charles Dobson – no financial interests to declare

Abstract

Introduction

In 2002, the National Institute for Clinical Excellence concluded that the multiple sclerosis disease modifying therapies; interferon-β and glatiramer acetate, were not cost-effective over the short term but recognised uncertainties over longer term modelling. The UK Risk-sharing Scheme (RSS) was thus set up to ensure cost-effective provision by measuring long-term disability progression in patients prescribed these therapies, and comparing with a modelled natural history cohort. It was agreed that the price of the drugs would be adjusted, if necessary, at each 2 year analysis to achieve the predefined cost per QALY target. The first 2 year analysis identified problems with the model, mainly focussed on unforeseen limitations of the chosen natural history database. This paper outlines the identification of a more suitable untreated cohort and the work undertaken to improve the Markov model.

Methods

All known international databases were screened to identify the most suitable comparator to the UK RSS cohort.

Using transition probabilities from the selected cohort, the original discrete Markov model was compared to a continuous model, with and without the addition of baseline covariates, looking for the best predictive model of the actual progression of the cohort from baseline data alone, assessed by "goodness-of-fit" analysis. .

Analysis

The British Columbia Multiple Sclerosis database was selected as most suitable for the scheme's purpose.

A continuous Markov model with "age at onset" as a binary covariate was deemed the most suitable model for future RSS analysis, providing the added benefit of allowing the use of data previously excluded due to time-window constraints.

Conclusion

A new statistical analysis plan has been developed for the UK RSS which will be used for price adjustment calculations for future RSS analyses. We believe this will provide a more valid and robust methodology upon which to base future decisions.

Strengths and limitations of this study

Strengths of this study:

- The validation of an analysis model for observational studies using natural history datasets as a comparator.
- The identification of an MS analysis model which can be applied over a 10 year follow up period.
- The identification of an analysis model which can use data collected at anytime point within the follow up period.
- The identification a model which uses data from a set of MS patients to predict outcomes in a different set of MS patients and in a different MS cohort.

Limitations of this study:

- This model cannot address unseen variations occurring due to the lack of randomisation.
- This model cannot address the bias in outcome due to lost to follow up patients
- Data from the natural history cohort of untreated patients was collected prior to that of the treated cohort and this model cannot adjust for any changes in the background outcome of untreated patients.
- Different techniques to assess long term effects such as propensity modelling cannot be directly compared to this methodology.

Introduction

In January 2002, the UK's National Institute of Clinical Excellence (NICE) opted not to recommend the use of the disease modifying therapies (DMTs) interferon-β and glatiramer acetate for multiple sclerosis (MS) on the basis of cost-effectiveness analyses using data derived from the pivotal 2-3 year randomised controlled trials.[1] However, they recognised that uncertainties over the assumptions made in the modelling could unpredictably influence the long-term estimates of cost effectiveness. Thus in February 2002 the UK's Department of Health launched the 'Risk-sharing Scheme' (RSS)[2] with a circular entitled the "Cost effective provision of disease modifying therapies for people with multiple sclerosis" in collaboration with the Association of British Neurologists (ABN), the MS Trust, the MS Society and the pharmaceutical companies manufacturing interferon-β and glatiramer acetate. Between 2002 and 2005 the scheme enrolled over 5,000 MS patients initiating a DMT in the UK, with the aim of measuring their disability annually over a ten year period.

The original cost effectiveness model[3] produced a target outcome based upon transition probabilities obtained from a pre-existing natural history (DMT naive) cohort of patients from London, Ontario, Canada along with hazard ratios from the pivotal randomised control trials (unpublished data provided to the Department of Health by the manufacturers). Complementary quality of life data collected by the MS Trust[4] and cost data from Kobelt et al[5] were used to populate the cost-effectiveness model. The targets ensured that the UK's National Health Service benchmark of £36,000 (46,000 Euro / 56,000 US dollars) per quality adjusted life year (QALY) was reached over a 20 year projection, based on a planned 10 year follow up period within the RSS with 2 yearly interim analyses. At the start of the scheme, the drug costs were reduced where necessary to ensure the predicted targets were on course to reach the 20 year cost effectiveness target.

The two year analysis revealed significant inconsistencies in a number of sensitivity analyses.[6] Depending on the underlying assumptions, some analyses suggested that observed disability progression in the treated cohort was worse than that predicted from the historical untreated cohort while others demonstrated the contrary effect. A detrimental effect of DMT did not match the described effect on short term, 2-3 year, disability seen in the randomised placebo controlled trials.[7-12] With the predetermined analytical approach (based on a discrete Markov model) appearing to produce unreliable results with wide variation, a decision was made to postpone any decision on cost

effectiveness to allow for a reappraisal of the process and to reconsider whether the statistical models and control data chosen were "fit for purpose."

In retrospect, both the control data set and analysis model selected, when setting up the RSS, were found to have intrinsic flaws that made them unsuitable for the task.[6] The natural history cohort (from London, Ontario, Canada) was unexpectedly found to contain retrospectively smoothed disability data (rather than actual, real-time collected disability scores), censoring any improvement in EDSS. Comparing our uncensored treated cohort to data retrospectively smoothed in this way would have the effect of unpredictably underestimating any treatment effect. In addition, individual-level patient data were not available from the London, Ontario cohort, which prevented precise baseline matching between the two cohorts, limiting our validation of the underlying (Markov) model for disease progression. Furthermore, there were only 342 patients matching the ABN prescribing criteria from which to generate the models.

This paper outlines the development of a more appropriate analysis plan and the choice of a cohort fit for the needs of the scheme. The method described will be applied in the 4 and 6 year cost-effectiveness analyses. The analysis plan was approved by the scheme's independent Scientific Advisory Group in December 2012 in advance of unlocking the newly collected 4 and 6 year UK Risk-sharing Scheme data planned for autumn 2013.

Methods

Identification of a new multiple sclerosis natural history dataset

An initial screen to identify all published natural history data sets was performed by reviewing the literature and consulting with international experts. Selection criteria included availability of Expanded Disability Status Scale (EDSS) score measurements and access to the unprocessed (actual) scores (i.e. no data smoothing or other data manipulation). Other factors considered were size of the database, prospective data collection and length of follow-up, and the broader setting such as a close match to the UK in terms of the health system and MS prevalence in the underlying population. The British Columbia Multiple Sclerosis (BCMS) database, Canada (est. 1980) was identified as the best natural history comparator cohort for our purposes.[13,14] In this dataset – as in the RSS – actual EDSS scores were recorded prospectively. It is estimated to capture 80% of the BC MS population[15,16] and

as such is considered representative of the wider MS population. EDSS scores were recorded by MS specialist neurologists after a face-to-face consultation with the patient; this typically occurred at the annual MS clinic visit. Patient data was not truncated if secondary-progressive MS was reached; i.e. all relapse-onset MS patients and their respective EDSS scores were considered eligible. By 2004, the database had records for over 5900 patients spanning 28 years (>25,000 cumulative years) of prospective follow-up. Until 1996 DMTs were not widely available in British Columbia.

Patient and data selection from the BCMS database.

In order to generate relevant data for our needs, patients were only selected from the BCMS database if they fulfilled the 2001 Association of British Neurologists (ABN) criteria for interferon-β and glatiramer acetate (IFN- β/GA) use (adapted from Appendix IV Health Service Circular 2002/004), defined as: EDSS≤6.5; ≥18 years old; two relapses in the last 2 calendar years.

Baseline for each patient was the 'first eligibility date,' meaning the first date at which a patient fulfilled the ABN eligibility criteria. Only patients with definite MS (Poser criteria[17]) and a minimum of two EDSS scores at least 9 months apart were considered.

In order to be comparable with the RSS data the following adjustments and selection were applied:

1) EDSS scores taken during a relapse or when disability was affected by other factors considered largely unrelated to MS (e.g. hip fracture) were excluded.

2) For the original discrete Markov model (see below) as well as visual presentation of the yearly descriptive data (see under results), annual EDSS scores were needed. However, as is typical in clinical practice, not all visits / EDSS assessments occurred at exactly yearly intervals and the exclusion of some EDSS scores (e.g. due to a relapse or hip fracture) also affected the availability of a yearly score. Therefore, data was selected such that only EDSS scores one year apart (+/- three calendar months) were considered. See appendix 1 for further details.

3) For the continuous Markov model, (see below) all eligible EDSS scores were used regardless of their measurement interval i.e. no yearly data selection, as in (2), was needed.

4) All patient data was truncated to the end of 1995 (i.e. the last full year in which the DMTs were not widely available in BC Although initially it was planned to truncate individual patient profiles only once a DMT was initiated (in order to maximise the number of EDSS assessments), even when this extended past 1995 when treatment would have been available). It became apparent that this introduced a bias into the data, likely related to 'indication bias,' whereby patients 'doing well' would be less likely to start a DMT.

Analysis

The primary purpose of the analytical work was to find the best statistical model able to predict EDSS progression in a natural history cohort based on entry demographic and clinical data. The following models were applied in the current study and their performances were critically evaluated.

a) The discrete Markov model[18] as in the original 2 year analysis[6] i.e. disability scores (EDSS) had to be measured at discrete, fixed time points.

b) A continuous Markov model allowing for EDSS scores to be collected at unevenly spaced time intervals, as is typical in clinical practice.[19] Such a model also allowed covariates to be included. This model allows for more complete use of EDSS scores collected at irregular time intervals both in the BCMS and RSS cohorts.

With regard to the "MS course" (i.e. relapsing remitting vs. secondary progressive) as a potential covariate, we did not distinguish between these disease states when developing the Markov models because secondary progressive MS is simply a later stage of the relapsing remitting form of the disease and the transition has considerable overlap.

To keep computations feasible, only integer EDSS values were used and fractional values rounded down (i.e. EDSS 1.5 was scored as 1, 2.5 was scored as 2 etc.). These were referred to as the ten EDSS 'states' (1-10). Transition probability and intensity matrices as the output of these models were then used to predict disease progression in terms of EDSS as follows.

Predicting outcomes in the continuous Markov model (b)

A multi-state model algorithm ('R' library 'msm'[19]) allows the EDSS distribution to be predicted at any time *t*. See appendix 2 for further details).

Covariates considered in the models

The selection of potential covariates by the scientific advisory group was based upon; (a) those which were reliably and consistently recorded in both the BCMS and the RSS database and (b) *a priori* knowledge of those associated with the outcome of disability progression. As a result, sex, age at MS symptom onset, as well as disease duration and disability (EDSS) - combined into a Multiple Sclerosis Severity Score (MSSS)[20] - were tested in the continuous Markov model with up to two covariates. In addition, for the more promising models an alternative model was considered with dichotomous covariates (split at the median) replacing the continuous variables. This has the advantage that the resulting model can be formulated as the aggregate of a small number of discrete Markov models, so computations can be carried out without requiring special software.

Critical evaluation of the models was performed using the following validation techniques, with the goal being to identify the most appropriate model to represent the natural progression of MS. See appendix 3 for further details.

1) Transition probabilities derived from the complete eligible, BCMS natural history data were applied to the baseline data to predict outcomes over the subsequent 10 years to assess how well it matched the observed data from which the model was derived.

2) The BCMS dataset was repeatedly randomly divided into two subsets of equal size, with one half only being used to derive transition probabilities (as in #1). The probabilities derived from this half were then applied to the baseline characteristics of the second half, generating a model whose goodness of fit could be judged against the actual ,observed 10 year disability data of this second half.

Measuring Goodness of Fit

Goodness of fit was assessed via visual inspection of the graphical displays as well as numerically. These included progression over time (mean EDSS profiles) for the cohort as a whole as well as comparisons with the proportions in a particular EDSS state over time.

For the numerical assessment a classical mean square prediction error (weighted root mean square over years of the prediction error in the average quantity shown, weighted by the number of patients contributing data in the given year) and the likelihood, resulting from the maximum likelihood algorithm were calculated for each of the covariate models to allow comparison.

This study was approved by the University of British Columbia's Clinical Research Ethics Board (H08-01544)'

Results

Data Description

The baseline demographics showed the BCMS and RSS cohorts to be remarkably well matched. Patient characteristics are shown in table 1.

The natural history BCMS comparator dataset comprised of 898 patient profiles with 7335 EDSS scores providing 6357 transitions between consecutive EDSS states, i.e 6357 'events' where EDSS values were recorded at consecutive visits. In any given "transition," a patient's EDSS could increase, decrease or stay the same

Discrete Markov model

When applying the discrete Markov model to the BCMS reference data, the goodness of fit was unsatisfactory, underestimating EDSS in earlier years and overestimating in later years (see figure 1). Consequently, the discrete Markov model was no longer considered appropriate, and development of a continuous Markov model was pursued.

Continuous Markov models

The following continuous 10 state Markov models (EDSS 0 to 9), with and without covariates, were evaluated:

1. Model without covariates
2. One covariate model with age at onset*
3. One covariate model with MSSS* at baseline
4. One covariate model with disease duration* at baseline
5. One covariate model with sex

6. Two covariate model: sex and age at onset*
7. Two covariate model: MSSS* at baseline and age at onset*
8. Two covariate model: disease duration* at baseline and age at onset*

- *two variants were implemented: continuous (original) data and a 'binary' version with the median used for categorisation.

There was a systematic deviation with overestimation when the continuous Markov models without covariates were applied (figure 2). Hence these models were not considered further.

After validation was repeated for all covariate models (table 2), it was noted that inclusion of a second covariate did not reveal any additional benefits. With one covariate, the model with "age at onset" as a binary covariate was selected because it displayed the smallest -2 log likelihood and minimal EDSS prediction error, see table 2. Further, the goodness of fit was acceptable when comparing the predicted and observed EDSS profiles, as shown in Figure 3a. A more detailed comparison of observed and expected proportions 'per EDSS state' is shown in Figure 3b which confirmed that no systematic deviations were present which might otherwise have been overlooked. It was concluded that only random fluctuation remained, and a systematic deviation was no longer visible. When comparing figure 1 with figures 2 or 3 it should be noted that the former is based on the *annual* EDSS data which were obtained as described in Appendix 1 while figures 2 and 3 show the EDSS at any time *t,* , i.e. not necessarily when an observation was recorded (for details on how to define and calculate what is the observed EDSS at a given time see Appendix 2).

Using this 'best' model, transition probabilities were extracted from half of the BCMS cohort and applied to the other half. This gave good predictions, with the mean EDSS profiles (observed versus predicted) being similar to each other and to those of the entire cohort.

Finally, further (external) validation was undertaken using a Welsh dataset of untreated MS patients collated by the Cardiff neurology team. When using the model with "age at onset" as the (only) binary covariate in a continuous Markov model we observed a pattern of congruence similar to that visible in figure 3, but limited to the comparatively shorter observation time in the Welsh cohort (data not shown.) This observation supported our choice of the 'best' model in the sense of finding an appropriate model for EDSS progression in untreated MS patients.

In summary, the continuous Markov model with a single covariate - onset age - was considered the model of choice to be used in future RSS analysis.

Discussion

This paper outlines the successful identification of a more suitable natural history cohort for the UK MS risk-sharing scheme, with the British Columbia, Canada dataset now replacing the London, Ontario, Canada cohort in the RSS analysis plan. The analytical work is based on a Markov model which has been frequently used for ordinal data from relapsing (remitting) diseases, especially MS.[21,22,23]

Further, because use of the British Columbian data has now allowed access to a richer dataset, including full access to original, 'real-time' disability (EDSS) assessments, as well as individual patient-level, we have been able to explore and develop more appropriate approaches. Specifically, we were able to employ more advanced statistical models, making use of all the available data and including clinically relevant patient-level characteristics as covariates in order to identify the most accurate predictive model to be applied to the RSS. Finally, we observed that to minimize 'indication bias' in relation to initiation of a DMT in the natural history cohort (British Columbia), censoring (data truncation) was more appropriate at the population (rather than individual) level.

Findings from our validation procedures indicate future feasibility with respect to obtaining reliable cost-effectiveness results in the upcoming 6 year RSS analyses. For instance, visualisation of the predicted and observed outcomes in the final model showed almost perfect overlap, with a one-covariate model, with no additional improvements from introducing further covariates. In addition, the final model was able to predict accurately the MS disease course (disability) in half of the cohort (randomly selected) having obtained the transition probabilities from the other half. We were also able to show that the model showed good fit when using the BCMS dataset to predict outcome of an untreated Welsh cohort. These observations along with the baseline comparability of the BCMS and the RSS cohorts suggest the transition probabilities from the BCMS cohort within this model can be used to predict the untreated progression of patients in the RSS.

An additional strength of this continuous model is the ability to include all valid disability (EDSS) assessments, regardless of their exact timing, maximising data usage. We acknowledge the potential limitations of using of a historical control from a geographically discrete population. It is possible that the natural history of MS has changed over time or that the BCMS population is not representative of a UK one. However, in British Columbia, it has been previously shown that disability progression (as measured by the EDSS) has not substantially changed overtime (1980-2009[24]). Further, we have previously shown that the use of a 'contemporary' untreated control cohort – i.e. where patients are potentially eligible for a DMT in an era when the DMTs are readily available, but remain untreated - is subject to indication bias and thus a historical control cohort, with data collected pre-DMT use, is likely to be more appropriate.[25]

Observational studies, such as the RSS, provide a pragmatic approach when assessing drug effectiveness in a disease such as MS. Because MS disability accrues over decades, the cost effectiveness of disease modifying treatments cannot be assessed by short-term randomised controlled trials. However, observational studies are not without their own unique challenges. Identifying and validating models to predict the untreated outcome of treated cohorts is a crucial step to measuring the long-term benefits of MS treatments. MS is the commonest cause of progressive disability in the western world, thus identification of treatments that might significantly impact long-term disability outcomes in MS could have major cost and quality of life benefits. Additionally, any models developed here would be readily transferable to other chronic diseases.

In summary, the current model described here will form the basis for calculating the drug cost per QALY and for informing decisions on price adjustment in order to deliver the treatments cost effectively to UK MS patients. Further work on repeated measures modelling, testing the models on other untreated appropriate MS datasets and identifying sensitivity analyses (such as the effect of drop outs, switching to a different class of DMT and the effects of treatments on backward transitions, i.e. disability improvements) are also planned.

All authors 1) made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data 2) were involved in drafting the article or revising it critically for important intellectual content and 3) were involved in the final approval of the version to be published. JP is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript. TB is the senior statistician in the scheme and the analysis of this work, was involved in the drafting and revision of this manuscript. HT, FZ and JO were responsible for the preparation of the BCMS dataset, analysis of the data, and drafting and revision of the manuscript, MB was a clinical lead for the UK risk sharing scheme, input into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript, MD is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the interpretation of the analysis and was involved in the drafting and revising of the manuscript, CD is a department of health advisor for the scheme, inputs into the scientific advisory panel, and was involved in the analysis of the data, its interpretation and was involved in the drafting and revising of the manuscript.

Acknowledgements

At the outset of the Risk-sharing Scheme agreement was reached between the UK Department of Health and the four manufacturers of the DMTs that all costs for this research project would be split 72 centres participated in the collection of data for the Risk-sharing Scheme from across the UK. Data was collected over a 10 year period and thus changes have been seen within the clinical teams. The authors are grateful to all the clinicians for the work undertaken at the centres which made the RSS possible. A list of the centres can be found at http://www.mstrust.org.uk/research/risksharingscheme/

The British Columbia MS (BCMS) database was created by Donald Paty FRCPC and Donald Studney MD FRCPC and maintained over the years by grants from various sources, including the MS Society of Canada, the MS/MRI Research Group, the US National MS Society and the Canadian Institutes of Health Research. It has essentially been contributed to by: Donald Paty, Stanley Hashimoto, John Hooge, Lorne Kastrukoff, Joel Oger, Anthony Traboulsee, Virginia Devonshire, Ana-Luiza Sayao. A full list of the consultant neurologists involved in the collection of data for the BCMS database can be found at http://www.mstrust.org.uk/research/risksharingscheme/
The authors are very grateful for access to the BCMS data.

The independent Scientific Advisory Group has been chaired by Prof Lilford and involved Prof Yoav Ben-shlomo, Prof Richard Gray and Prof Pelham Barton.

The MS Trust has performed an administrative coordinating role in the Risk-sharing Scheme and PAREXEL have performed the data collection from the sites and undertaken the analytical work.

The authors are grateful to Prof Neil Robertson and his team (Cardiff University) and Prof Kate Tilling, (Dept of Social Medicine, Bristol University) for the availability of the Welsh dataset and its preparation for our analysis.

References

1. National Institute for Clinical Excellence. Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. NICE Technology Appraisal Guidance No. 32. London: NICE; 2002.

2. Department of Health. Cost effective provision of disease modifying therapies for people with multiple sclerosis. London: Health Service Circular 2002/004 London: Stationery Office; 2002.

3. Chilcott J, McCabe C, Tappenden P, O'Hagan A, Cooper NJ, Abrams K, et al. Modelling the cost-effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. BMJ 2003;326:522.

4. Hemmett L, Holmes J, Barnes M, Russell N. What drives quality of life in multiple sclerosis? QJM. 2004;97:671-6.

5. Kobelt G, Lindgren P, Parkin D, Francis DA, Johnson M, Bates D, et al. Costs and quality of life in multiple sclerosis. A cross-sectional observational study in the UK. Scandinavian Working Papers in Economics, 2000 (http://swopec.hhs.se/hastef/papers/hastef0398.pdf ; date accessed: 28-Aug-2013)

6. Boggild M, Palace J, Barton P, Ben-Shlomo Y, Bregenzer T, Dobson C, et al.. Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. BMJ 2009;339:b4677.

7. The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. Neurology. 1993;43:655-61

8. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Ann Neurol. 1996;39:285-94.

9. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. Lancet. 1998;352:1498-504.

10. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. Neurology. 1995;45:1268-76.

11. European Study Group on interferon beta-1b in secondary progressive MS. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. Lancet. 1998;352:1491-7.

12. Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-Beta-1a in MS (SPECTRIMS) Study Group. Randomized controlled trial of interferon- beta-1a in secondary progressive MS: Clinical results. Neurology. 2001;56:1496-504.

13. Tremlett H, Paty D, Devonshire V. Disability progression in MS is slower than previously reported. Neurology 2006;66:172-7.

14. Tremlett H, Zhao Y, Rieckmann P, Hutchinson M. New perspectives in the natural history of multiple sclerosis. Neurology 2010;74:2004-15.

15. Sadovnick AD, Ebers GC, Wilson RW, Paty DW. Life expectancy in patients attending multiple sclerosis clinics. Neurology. 1992;42:991-4.

16. Sweeney VP, Sadovnick AD, Brandejs V. Prevalence of multiple sclerosis in British Columbia. Can J Neurol Sci. 1986;13:47-51.

17. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol. 1983;13:227-31.

18. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Econ. 2002;11:33-42.

19. Jackson CH, Sharples LS, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. Journal of the Royal Statistical Society: Series D (The Statistician) 2003;52:193-209.

20. Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology. 2005;64:1144-51.

21. Albert P. A Markov model for sequences of ordinal data from a relapsing-remitting disease. Biometrics 1994;50:51-60.

22. Bell C, Graham J, Earnshaw S, Oleen-Burkey M, Castelli-Haley J, Johnson K., Cost-effectiveness of four immunomodulatory therapies for relapsing-remitting multiple sclerosis: a Markov model based on long-term clinical data. J Manag Care Pharm 2007;13:245-61.

23. Mandel M., Betensky R., Estimating time-to-event from longitudinal ordinal data using random-effects Markov models: application to multiple sclerosis progression. Biostatistics 2008;9:750-64.

24. Shirani A, Zhao Y, Kingwell E, Rieckmann P, Tremlett H. Temporal trends of disability progression in multiple sclerosis: findings from British Columbia, Canada (1975-2009). Mult Scler. 2012;18:442-50.

25. Shirani A, Zhao Y, Karim ME, Evans C, Kingwell E, van der Kop ML, et al. Association between use of interferon beta and progression of disability in patients with relapsing-remitting multiple sclerosis. JAMA. 2012;308:247-56.

.

Table 1: Characteristics of patients reaching the Association of British Neurologists criteria in the British Columbia MS database after 1980 (the 'natural history' untreated comparator cohort) and the Risk-sharing Scheme cohort.

Baseline (eligible for DMT)	BCMS (1980-1995*)	RSS full cohort RSS analysis cohort
N	898	5610 4138
Females: n (%)	666 (74.2%)	4162 (74.2%) 3125 (75.5%)
Age at baseline, years: mean (SD; range) [years]	37.2 (9.32; 18 - 69)	39.4 (9.05; 15 - 73) 38.4 (8.58; 18 - 73)
Age at onset of MS, years: mean (SD; range)	29.3 (8.65; 3 - 61)	30.5 (8.52; 5 - 68) 30.5 (8.38; 5 - 68)
Disease duration at baseline, years: mean (SD; range) [years]	7.9 (6.89; 0.2 to 38.9)	8.8 (7.47; 0 - 46) 7.7 (6.62; 0 - 41)
SPMS documented at baseline# n (%)	141 (15.7%)	772 (13.8%) -
Relapses in the last two-years prior to eligibility: median (quartiles)	2 (2 - 3)	3 (2 - 3) 3 (2 - 3)
First eligible EDSS: median (quartiles; range)	2 (1, 3.5; 0-6.5)	3.5 (2.0, 5.0; 0 - 8.0) 3.0 (2.0, 4.0; 0 - 6.5)

Key: *'Eligibility' refers to the first time a patient fulfilled the ABN criteria*data was truncated to 1995 in the final models to minimize DMT exposure in the cohort*

#all were still DMT eligible

RSS=Risk-sharing Scheme ; BCMS=British Columbia MS database; SD=standard deviation; EDSS=Expanded disability status score ; DMT=disease modifying treatment

Table 2: "Goodness of fit" statistics for the ten state* disability (EDSS) Markov models

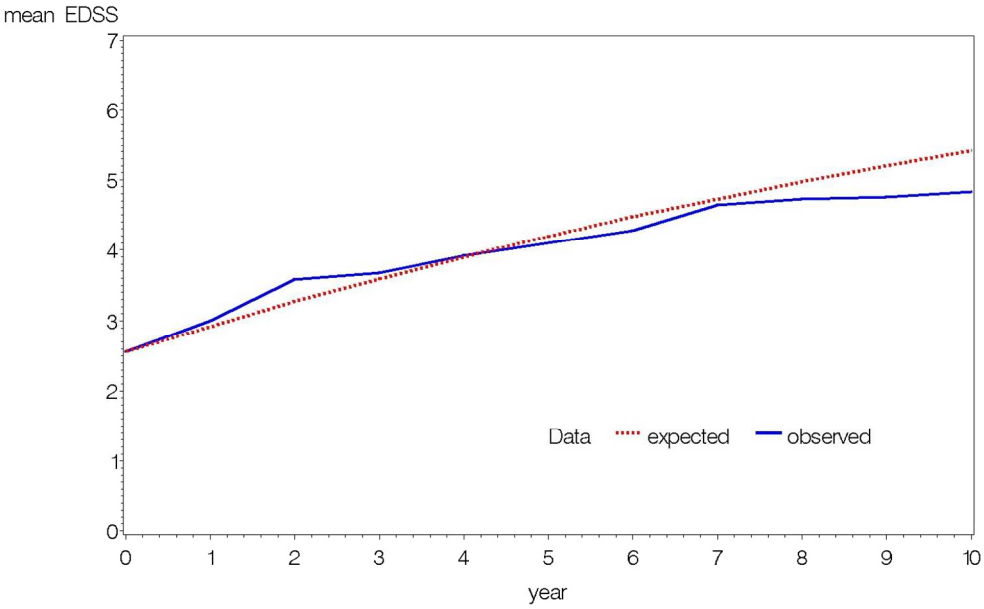
Description of each ten-state[1] disability model	Minus 2 log likelihood[2] x 1,000	Prediction errors 1-10)[3] Cells	EDSS	(years Utility
No covariates	17.152	2.20	0.24	0.022
One covariate models				
Age at onset, binary	17.458	1.39	0.09	0.009
Age at onset, continuous	17.599	1.58	0.13	0.007
MSSS at baseline, binary	17.460	1.41	0.10	0.008
MSSS at baseline, continuous	17.457			
Disease duration, binary	17.462	1.33	0.10	0.009
Disease duration, continuous	17.557			
Sex	17.470	1.32	0.10	0.008
Two covariates models				
Sex and age at onset, binary	17.603	1.51	0.14	0.007
Sex and age at onset, continuous	17.618			
Age at onset and MSSS, binary	17.609	1.53	0.14	0.007
Age at onset and MSSS, continuous	17.618			
Age at onset and disease duration, binary	17.603	1.52	0.14	0.007
Age at onset and disease duration, continuous	17.618			

[1]the ten disability states refer to EDSS 0 to 9, i.e. EDSS 0 is "state 1", EDSS 1 is "state 2" etc.

[2]log likelihood statistic as calculated by 'msm' module, see Jackson[19] for details; lower values implying a better model (to be compared within each class of models, e.g. one-covariate and two-covariate models)

EDSS= Expanded disability status score ; MSSS= Multiple sclerosis status score

[3]Prediction errors, averaged over years 1-10, for (a) the EDSS distribution in individual cells, (b) average EDSS, (c) average utility (see definitions in the appendix 3, comparing the values predicted by the model with the "observed" values using the method of midpoint interpolation (see appendix 2).



Transition probabilities obtained from the BCC dataset using the discrete Markov model were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted mean EDSS outcome (red) and compared to the observed mean EDSS course of the cohort (blue).
402x264mm (96 x 96 DPI)

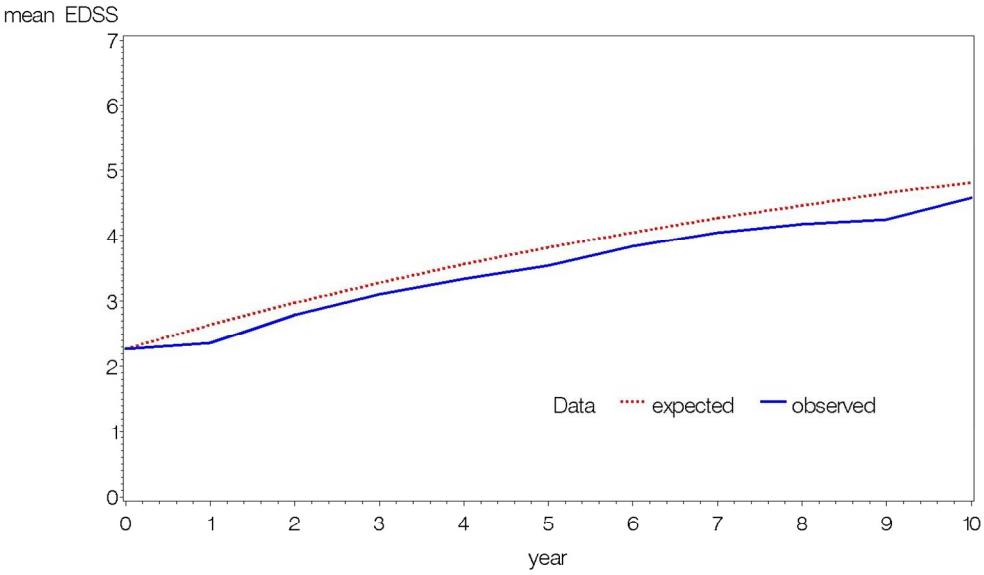


Figure 2: Transition probabilities obtained from the BCC dataset using the continuous Markov model were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted mean EDSS outcome (red) and compared to the observed mean EDSS course of the cohort (blue).
402x264mm (96 x 96 DPI)

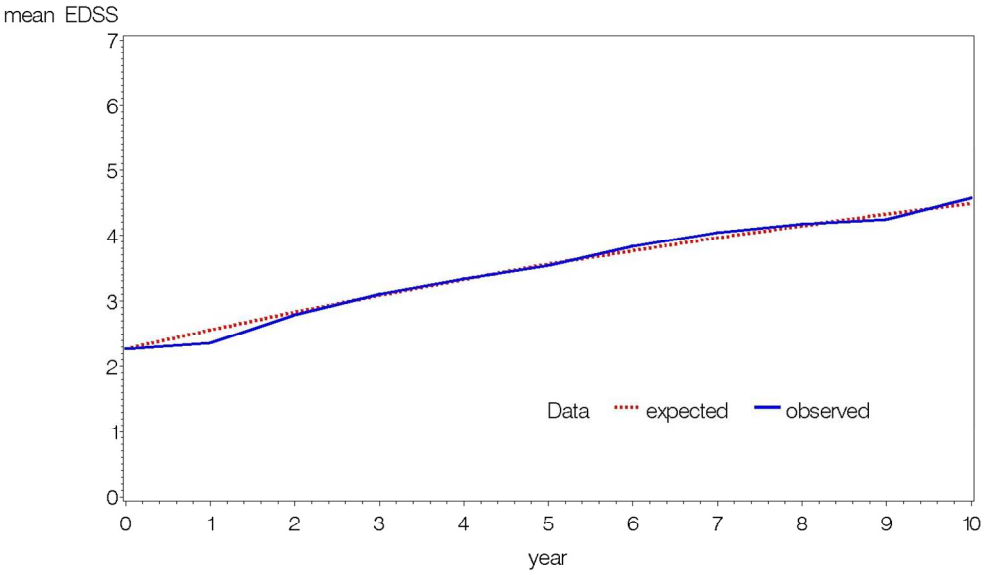
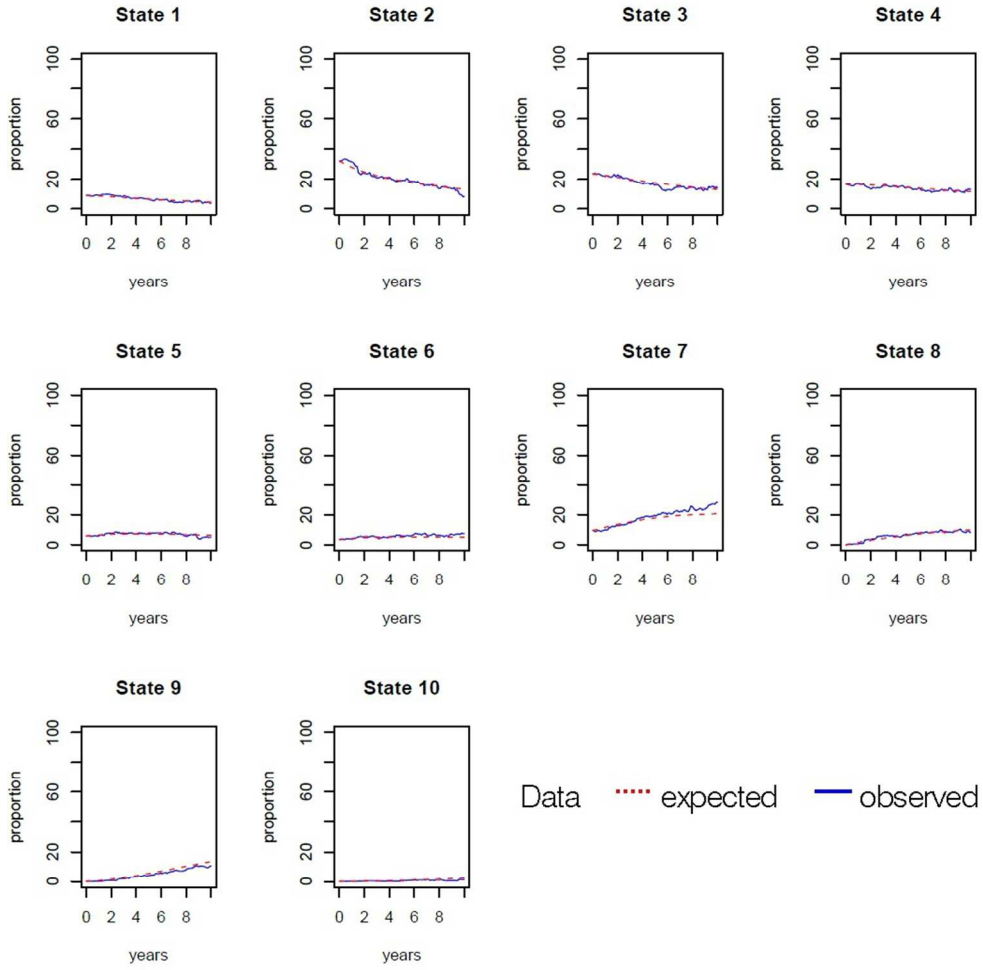


Figure 3: Transition probabilities obtained from the BCC dataset using the continuous Markov model with one covariate "age at onset" (binary version) were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted outcome (red) compared to the observed course of the cohort (blue):
a) Mean EDSS shown in the predicted and actual cohorts
b)The proportion of patients predicted to be in each of the 10 EDSS states over time (state 1; EDSS 0, state 2; EDSS 1 and 1.5, state 3; EDSS 2.0 and 2.5, state 4; EDSS 3.0 and 3.5, state 5; EDSS 4.0 and 4.5, state 6; EDSS 5.0 and 5.5, state 7; EDSS 6.0 and 6.5, state 8; EDSS 7.0 and 7.5, state 9; EDSS 8.0 and 8.5, state 10; EDSS 9.0 and 9.5).

402x264mm (96 x 96 DPI)



283x280mm (96 x 96 DPI)

Appendix 1 EDSS data extraction from the BCMS database

For the discrete Markov model and for tabular display of annual data the EDSS data were extracted and processed as follows. If the baseline EDSS fell when it was not the 'usual' yearly visit period, then subsequent EDSS scores could be lost. To maximize the number of EDSS transitions per patient, but keeping the yearly (+/- three calendar months) interval, EDSS scores were also 'individualized' as follows: the baseline EDSS became the baseline year, within that year, we searched to find the optimal new, individualized baseline for which the patient would, over the coming years have the most number of yearly EDSS scores. For some patients, this new individualized baseline date would not coincide with a clinic visit and would therefore not have an EDSS score present. When a baseline EDSS was required (e.g. for the discrete Markov model), data was also 'shifted' such that each patient would have a baseline EDSS. This lag between baseline ('eligibility') and EDSS was considered consistent with clinical practice in that it is not unusual to have a lag time between a patient becoming eligible for treatment and actual treatment initiation.

Appendix 2 Algorithms to forecast EDSS distributions at any given time

The 'msm' algorithm allows the EDSS distribution to be forecasted at any time t.19 To define what is the actual EDSS at a given time t (i.e. not necessarily when an observation was recorded) 'msm' offered two variants: (i) the last observation carried forward (LOCF) for each individual patient, and (ii) a 'midpoint interpolation' algorithm in which the EDSS state for a given patient at time t was taken as the score closest in time to the actually observed EDSS. Suppose an individual was observed in EDSS states Sr-1 and Sr at two consecutive times tr-1 and tr, and we wanted to estimate ''observed' proportions at a time t between tr-1 and tr. LOCF then meant that individuals were assumed to be in state Sr-1 at time t, the same state as they were at tr-1. Midpoint interpolation meant if $t \leq (tr\text{-}1 + tr) / 2$, the midpoint of tr-1 and tr, the state at t was assumed to be Sr-1,, otherwise Sr. Option (i) would be more appropriate if EDSS values were always measured immediately after each transition. Option (ii) would be more appropriate if EDSS values were measured at mixed time intervals ('fixed or random'). Option (ii) was considered to mimic the clinical setting more closely and is also appears more applicable in progressive diseases.19 Therefore all results presented in this paper were based on this 'midpoint interpolation' approach.

For the continuous Markov model we limited the range of transitions which can be regarded as 'instantaneous' to +/- 3, which meant that at any time t an instantaneous progression (or improvement) into another EDSS state was only possible when not exceeding three consecutive states (an instantaneous transition from 1 to 2, 3, or 4 was possible, for example, whereas 1 to 5 was not). This restriction was recommended to avoid computationally inefficient modelling of hazard rates which were virtually 0.

Appendix 3 Validation of the models

The following validation techniques were applied when evaluating the different Markov models.

The most straightforward method for the discrete Markov model consisted of applying the transition matrix (and the 2nd, 3rd etc. power) to the vector of the baseline EDSS distribution in the BCMS reference database, calculating the forecasted EDSS distribution for t=1, 2, 3… years and comparing against the actual EDSS. A similar validation was performed in the continuous Markov model, as described by Jackson.1.

As a second validation we divided the BCMS dataset randomly into two subsets of equal size, using one half to derive the model separately for the two subgroups, and then assessing the goodness of fit in the other half of the dataset. Although Jackson19 emphasises that "Assessing the goodness of fit of this class of models […] is worth further research" we decided to use a classical mean square prediction error (weighted root mean square over years of the prediction error in the average quantity shown, weighted by the number of patients contributing data in the given year) to compare competing models. Moreover, the computed likelihood itself as a result of the maximum likelihood algorithm was used to rank the different one and two covariate models.

UK Multiple Sclerosis Risk-sharing Scheme: a new natural history dataset and an improved Markov model

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004073.R1
Article Type:	Research
Date Submitted by the Author:	10-Dec-2013
Complete List of Authors:	Palace, Jacqueline; John Radcliffe Hospital, Clinical Neurology Bregenzer, Thomas; PAREXEL International, Biostatistics Unit Tremlett, Helen; University Of British Columbia, Medicine (Neurology) Oger, Joel; University Of British Columbia, Division of Neurology Zhu, Feng; University Of British Columbia, Medicine (Neurology) Boggild, Mike; The Townsville Hospital, Neurology Department Duddy, Martin; The Newcastle upon Tyne Hospitals, Neurology Department Dobson, Charles; Department of Health,
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Health economics, Pharmacology and therapeutics, Research methods
Keywords:	Multiple sclerosis < NEUROLOGY, quality of life, risk sharing scheme, interferon-beta, glatiramer acetate, Markov model

UK Multiple Sclerosis Risk-sharing Scheme: a new natural history dataset and an improved Markov model

Palace, Jacqueline, Bregenzer, Thomas, Tremlett, Helen, Oge, Joel, Zhu, Fheng,Boggild, Mike, Duddy,Martin and Dobson, Charles

Palace, Jacqueline

John Radcliffe Hospital - Clinical Neurology
Headington , Oxford OX39DU
United Kingdom

Bregenzer, Thomas

PAREXEL International - Biostatistics Unit
Berlin
German

Tremlett, Helen

University Of British Columbia - Medicine (Neurology)
Vancover, British Columbia
Canada

Oge, Joel

University Of British Columbia - Division of Neurology
Vancover, British Columbia
Canada

Zhu, Fheng University Of British Columbia - Medicine (Neurology)
Vancover, British Columbia
Canada

Boggild, Mike

The Townsville Hospital - Neurology Department
Townsville, Queensland
Australia

Duddy,Martin

The Newcastle upon Tyne Hospitals - Neurology Department
Newcastle
United Kingdom

Dobson, Charles

Department of Health
Leeds
United Kingdom

The lead author* affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Abstract

Objectives: In 2002, the UK's National Institute for Health and Care Excellence concluded that the multiple sclerosis (MS) disease modifying therapies; interferon-β and glatiramer acetate, were not cost-effective over the short term but recognised that reducing disability over the longer-term might dramatically improve the cost -effectiveness. The UK Risk-sharing Scheme (RSS) was established to ensure cost-effective provision by prospectively collecting disability-related data from UK treated MS patients and comparing findings to a natural history (untreated) cohort. However, deficiencies were found in the originally selected untreated cohort and the resulting analytical approach. This study aims to identify a more suitable natural history comparator cohort and to develop a robust analytical approach using the new cohort

Design: The Scientific Advisory Group review, recommended the British Columbia Multiple Sclerosis (BCMS) database, Canada, as providing a more suitable natural history comparator cohort. Transition probabilities were derived and different Markov models (discrete and continuous) with and without baseline covariates were applied.

Setting: MS clinics and analysis groups in Canada and the UK.

Participants: From the BCMS database, 898 'untreated' MS patients considered eligible for drug treatment based on the UK's Association of British Neurologists criteria.

Outcome measure: The predicted disability, as measured by the Expanded Disability Status Scale (EDSS) score was collected and assessed for goodness of fit when compared to actual outcome.

Results: The BCMS untreated cohort contributed 7335 EDSS scores over a median 6.4 years (6357 "transitions" where EDSS values were recorded at consecutive visits) during the study period (1980-1995). A continuous Markov model with "age at onset" as a binary covariate was deemed the most suitable model for future RSS analysis.

Conclusion: A new untreated MS cohort from British Columbia, Canada has been selected and will be modelled using a continuous Markov model with age as a baseline covariate. This approach will now be applied to the treated UK RSS MS cohort for future price adjustment calculations.

Strengths and limitations of this study

Strengths of this study:

- Identification of a new natural history cohort for the UKRSS, consisting of untreated MS patients in an era when disease-modifying drugs for MS were not available, minimizing potential selection bias
- Identification and validation of a Markov model for disease progression in MS which can be applied to data collected in clinical practice over multiple years of follow up.
- The identification of an analytical model which can use data collected at any time point within the follow up period.

Limitations of this study:

- The study related to observational data collected in clinical practice; unseen or unmeasured confounding cannot be adjusted for.
- Different techniques to assess effectiveness of drugs in observational studies such as matching on propensity scores cannot be directly compared to this methodology.

Introduction

In January 2002, the UK's National Institute of Health and Care Excellence (NICE) opted not to recommend the use of the disease modifying therapies (DMTs) interferon-β and glatiramer acetate for multiple sclerosis (MS) on the basis of cost-effectiveness analyses using data derived from the pivotal 2-3 year randomised controlled trials.[1] However, they recognised that uncertainties over the assumptions made in the modelling could unpredictably influence the long-term estimates of cost effectiveness. Thus in February 2002 the UK's Department of Health launched the 'Risk-sharing Scheme' (RSS)[2] with a circular entitled the "Cost effective provision of disease modifying therapies for people with multiple sclerosis" in collaboration with the Association of British Neurologists (ABN), the MS Trust, the MS Society and the pharmaceutical companies manufacturing interferon-β and glatiramer acetate. Between 2002 and 2005 the scheme enrolled over 5,000 MS patients initiating a DMT in the UK, with the aim of measuring their disability annually over a ten year period.

The principle In the RSS is to use a Markov model to predict, for each DMT separately, the expected movement of patients between the EDSS states both "on" and "off" treatment. For patients "off" treatment, the model uses a matrix of transition probabilities derived from the actual progressions seen in the 'natural history' comparator cohort. These transition matrices are modified for patients "on" treatment by multiplying by the hazard ratio (relative rate of disease progression) derived separately for each DMT from the pivotal randomized controlled clinical trials. The model then predicts how the distribution of patients will evolve over a 20-year horizon, starting with the actual distribution at baseline for the primary analysis RSS cohort. Comparing the average observed loss of utility (average utility-weighted disease progression) for patients in the RSS to the expected loss calculated by the Markov model for patients "on" treatment; it is calculated as follows. The *expected* 'benefit' of treatment (with a specific DMT) is the "hypothetical" difference between the expected outcome without treatment and with treatment, as calculated in each case from the Markov model. The *actual* 'benefit' of treatment is the "observed" difference between the expected outcome without treatment and the actual outcome with treatment. The 'deviation' of the actual benefit from the expected is the primary outcome measure and calculated as a percentage of the expected benefit. This measure can have negative or positive values so that a negative deviation implies that the observed benefit was greater than predicted, a positive deviation suggesting that it was worse than

predicted and a value of 0 indicating that it was exactly as predicted. A positive deviation beyond the level agreed (confidential and individual between each pharmaceutical company and the Department of Health) would lead to a price adjustment down to achieve the target cost effectiveness. Details can be found in the Health Service Circular[2].

The original cost effectiveness model[3] produced a target outcome based upon transition probabilities obtained from a pre-existing natural history (DMT naive) cohort of patients from London, Ontario, Canada along with hazard ratios from the pivotal randomised control trials (unpublished data provided to the Department of Health by the manufacturers). Complementary quality of life data collected by the MS Trust[4] and cost data from Kobelt et al[5] were used to populate the cost-effectiveness model. The targets ensured that the UK's National Health Service benchmark of £36,000 (46,000 Euro / 56,000 US dollars) per quality adjusted life year (QALY) was reached over a 20 year projection, based on a planned 10 year follow up period within the RSS with 2 yearly interim analyses. Before being allowed to enter the scheme, the costs of each drug was assessed against the NICE bench mark over a 20 year time horizon. Price reductions were implemented to ensure each product reached the target cost per QALY using the original NICE calculations[3], an average 13.7% price reduction was achieved for the NHS at the outset of the Scheme.

The two year analysis revealed significant inconsistencies in a number of sensitivity analyses.[6] Depending on the underlying assumptions, some analyses suggested that observed disability progression in the treated cohort was worse than that predicted from the historical untreated cohort while others demonstrated the contrary effect. A detrimental effect of DMT did not match the described effect on short term, 2-3 year, disability seen in the randomised placebo controlled trials.[7-12] With the predetermined analytical approach (based on a discrete Markov model) appearing to produce unreliable results with wide variation, a decision was made to postpone any decision on cost effectiveness to allow for a reappraisal of the process and to reconsider whether the statistical models and control data chosen were "fit for purpose."

In retrospect, both the control data set and analysis model selected, when setting up the RSS, were found to have intrinsic flaws that made them unsuitable for the task.[6] The natural history cohort (from London, Ontario, Canada) was unexpectedly found to contain retrospectively smoothed disability data (rather than actual, real-time collected disability scores), censoring any improvement in EDSS.

Comparing our uncensored treated cohort to data retrospectively smoothed in this way would have the effect of unpredictably underestimating any treatment effect. In addition, individual-level patient data were not available from the London, Ontario cohort, which prevented precise baseline matching between the two cohorts, limiting our validation of the underlying (Markov) model for disease progression. Furthermore, there were only 342 patients matching the ABN prescribing criteria from which to generate the models.

This paper outlines the development of a more appropriate analysis plan and the choice of a cohort fit for the needs of the scheme. The method described will be applied in the 4 and 6 year cost-effectiveness analyses. The analysis plan was approved by the scheme's independent Scientific Advisory Group in December 2012 in advance of unlocking the newly collected 4 and 6 year UK Risk-sharing Scheme data planned for autumn 2013.

Methods

Identification of a new multiple sclerosis natural history dataset

The Scientific Advisory Group undertook a detailed examination of all the available dataset through literature reviews, expert opinion, discussion with the clinical leads for the RSS and discussion with the Sylvia Lawry Centre for Multiple Sclerosis Research, Germany (http://www.slcmsr.net). Selection criteria included availability of Expanded Disability Status Scale (EDSS) score measurements and access to the unprocessed (actual) scores (i.e. no data smoothing or other data manipulation). Other factors considered were size of the database, prospective data collection and length of follow-up, and the broader setting such as a close match to the UK in terms of the health system and MS prevalence in the underlying population. Whilst no single perfect dataset existed the British Columbia Multiple Sclerosis (BCMS) database, Canada (est. 1980) was identified as the best natural history comparator cohort for our purposes.[13,14] In this dataset – as in the RSS – actual EDSS scores were recorded prospectively. It is estimated to capture 80% of the BC MS population[15,16] and as such is considered representative of the wider MS population. EDSS scores were recorded by MS specialist neurologists after a face-to-face consultation with the patient; this typically occurred at the annual MS clinic visit. Patient data was not truncated if secondary-progressive MS was reached; i.e. all relapse-onset MS patients and their respective EDSS scores were considered eligible. By 2004, the database had

records for over 5900 patients spanning 28 years (>25,000 cumulative years) of prospective follow-up. Until 1996 DMTs were not widely available in British Columbia.

Patient and data selection from the BCMS database.

In order to generate relevant data for our needs, patients were only selected from the BCMS database if they fulfilled the 2001 Association of British Neurologists (ABN) criteria for interferon-β and glatiramer acetate (IFN- β/GA) use (adapted from Appendix IV Health Service Circular 2002/004), defined as: EDSS≤6.5; ≥18 years old; two relapses in the last 2 calendar years.

Baseline for each patient was the 'first eligibility date,' meaning the first date at which a patient fulfilled the ABN eligibility criteria. Only patients with definite MS (Poser criteria[17]) and a minimum of two EDSS scores at least 9 months apart were considered.

In order to be comparable with the RSS data the following adjustments and selection were applied:

1) EDSS scores taken during a relapse or when disability was affected by other factors considered largely unrelated to MS (e.g. hip fracture) were excluded.

2) For the original discrete Markov model (see below) as well as visual presentation of the yearly descriptive data (see under results), annual EDSS scores were needed. However, as is typical in clinical practice, not all visits / EDSS assessments occurred at exactly yearly intervals and the exclusion of some EDSS scores (e.g. due to a relapse or hip fracture) also affected the availability of a yearly score. Therefore, data was selected such that only EDSS scores one year apart (+/- three calendar months) were considered. See appendix 1 for further details.

3) For the continuous Markov model, (see below) all eligible EDSS scores were used regardless of their measurement interval i.e. no yearly data selection, as in (2), was needed.

4) All patient data was truncated to the end of 1995 (i.e. the last full year in which the DMTs were not widely available in BC Although initially it was planned to truncate individual patient profiles only once a DMT was initiated (in order to maximise the number of EDSS assessments), even when this extended past 1995 when treatment would have been available). It became apparent that this

introduced a bias into the data, likely related to 'indication bias,' whereby patients 'doing well' would be less likely to start a DMT.

Analysis

The primary purpose of the analytical work was to find the best statistical model able to predict EDSS progression in a natural history cohort based on entry demographic and clinical data. The following models were applied in the current study and their performances were critically evaluated.

a) The discrete Markov model[18] as in the original 2 year analysis[6] i.e. disability scores (EDSS) had to be measured at discrete, fixed time points.

b) A continuous Markov model allowing for EDSS scores to be collected at any time, i.e. at any unevenly spaced time intervals, as is typical in clinical practice.[19] Such a model also allowed covariates to be included. This model allows for more complete use of EDSS scores collected at irregular time intervals both in the BCMS and RSS cohorts.

With regard to the "MS course" (i.e. relapsing remitting vs. secondary progressive) as a potential covariate, we did not distinguish between these disease states when developing the Markov models because secondary progressive MS is simply a later stage of the relapsing remitting form of the disease and the transition has considerable overlap.

Predicting outcomes in the continuous Markov model (b)

A multi-state model algorithm ('R' library 'msm'[19]) allows the EDSS distribution to be predicted at any time *t*. See appendix 2 for further details).

To keep computations feasible, only integer EDSS values were used and fractional values rounded down (i.e. EDSS 1.5 was scored as 1, 2.5 was scored as 2 etc.). Moreover, 'msm' as a tool for *multi-state modelling* requires a consecutive numbering of (disease) states, starting with "1". Therefore the (rounded down) EDSS 0 became 'state 1', EDSS 1 'state 2' etc., leading to the ten EDSS 'states' (1–10) representing EDSS 0 –9. Transition probability and intensity matrices as the output of these models were then used to predict disease progression in terms of EDSS as follows.

Covariates considered in the models

The selection of potential covariates by the scientific advisory group was based upon; (a) those which were reliably and consistently recorded in both the BCMS and the RSS database and (b) *a priori* knowledge of those associated with the outcome of disability progression. As a result, sex, age at MS symptom onset, as well as disease duration and disability (EDSS) – combined into a Multiple Sclerosis Severity Score (MSSS)[20] were tested in the continuous Markov model with up to two covariates. In addition, for the more promising models an alternative model was considered with dichotomous covariates (split at the median) replacing the continuous variables. This has the advantage that the resulting model can be formulated as the aggregate of a small number of discrete Markov models, so computations can be carried out without requiring special software, especially since the EDSS values in the RSS have been collected at strict yearly intervals, as opposed to the BCMS data which was based on routine clinical practice, and therefore do not necessitate a continuous model.

Critical evaluation of the models was performed using the following validation techniques, with the goal being to identify the most appropriate model to represent the natural progression of MS. See appendix 3 for further details.

1) Transition probabilities derived from the complete eligible, BCMS natural history data were applied to the baseline data to predict outcomes over the subsequent 10 years to assess how well it matched the observed data from which the model was derived.

2) The BCMS dataset was repeatedly randomly divided into two subsets of equal size, with one half only being used to derive transition probabilities (as in #1). The probabilities derived from this half were then applied to the baseline characteristics of the second half, generating a model whose goodness of fit could be judged against the actual ,observed 10 year disability data of this second half.

Measuring Goodness of Fit

Goodness of fit was assessed via visual inspection of the graphical displays as well as numerically. These included progression over time (mean EDSS profiles) for the cohort as a whole as well as comparisons with the proportions in a particular EDSS state over time.

For the numerical assessment a classical mean square prediction error (weighted root mean square over years of the prediction error in the average quantity shown, weighted by the number of patients contributing data in the given year) and the likelihood, resulting from the maximum likelihood algorithm were calculated for each of the covariate models to allow comparison.

This study was approved by the University of British Columbia's Clinical Research Ethics Board (H08-01544)'

Results

Data Description

The baseline demographics showed the BCMS and RSS cohorts to be remarkably well matched. Patient characteristics are shown in table 1.

The natural history BCMS comparator dataset comprised of 898 patient profiles with 7335 EDSS scores providing 6357 transitions between consecutive EDSS states, i.e 6357 'events' where EDSS values were recorded at consecutive visits. In any given "transition," a patient's EDSS could increase, decrease or stay the same

Discrete Markov model

When applying the discrete Markov model to the BCMS reference data, the goodness of fit was unsatisfactory, underestimating EDSS in earlier years and overestimating in later years (see figure 1). Consequently, the discrete Markov model was no longer considered appropriate, and development of a continuous Markov model was pursued.

Continuous Markov models

The following continuous 10 state Markov models (EDSS 0 to 9), with and without covariates, were evaluated:

1. Model without covariates

2. One covariate model with age at onset*
3. One covariate model with MSSS* at baseline
4. One covariate model with disease duration* at baseline
5. One covariate model with sex
6. Two covariate model: sex and age at onset*
7. Two covariate model: MSSS* at baseline and age at onset*
8. Two covariate model: disease duration* at baseline and age at onset*

- *two variants were implemented: continuous (original) data and a 'binary' version with the median used for categorisation.

There was a systematic deviation with a slight overestimation when the continuous Markov model without covariates was applied (figure 2). Hence this model was not considered further.

After validation was repeated for all covariate models (table 2), it was noted that inclusion of a second covariate did not reveal any additional benefits. With one covariate, the model with "age at onset" as a binary covariate (median: 28 years) was selected because it displayed the smallest -2 log likelihood and minimal EDSS prediction error, see table 2. Further, the goodness of fit was acceptable when comparing the predicted and observed EDSS profiles, as shown in Figure 3a. A more detailed comparison of observed and expected proportions 'per EDSS state' is shown in Figure 3b which confirmed that no systematic deviations were present which might otherwise have been cancelled out when only looking at an average EDSS profile. It was concluded that only random fluctuation remained, and a systematic deviation was no longer visible. When comparing figure 1 with figures 2 or 3 it should be noted that the former is based on the *annual* EDSS data which were obtained as described in Appendix 1 while figures 2 and 3 show the EDSS at any time *t*, i.e. not necessarily when an observation was recorded (while the continuous Markov model takes into account all observations at any time *t* it is not straightforward to define what the 'observed EDSS' at any time *t* is in a graphic representation; for details on how to define and calculate what is the observed EDSS at a given time see Appendix 2).

Using this 'best' model, transition probabilities were extracted from half of the BCMS cohort and applied to the other half. This gave good predictions, with the mean EDSS profiles (observed versus predicted) being similar to each other and to those of the entire cohort.

In summary, the continuous Markov model with a single covariate - onset age - was considered the model of choice to be used in future RSS analysis. The respective transition matrices are presented in Table 3.

Discussion

This paper outlines the successful identification of a more suitable natural history cohort for the UK MS risk-sharing scheme, with the British Columbia, Canada dataset now replacing the London, Ontario, Canada cohort in the RSS analysis plan. The analytical work is based on a Markov model which has been frequently used for ordinal data from relapsing (remitting) diseases, especially MS.[21,22,23]

Further, because use of the British Columbian data has now allowed access to a richer dataset, including full access to original, 'real-time' disability (EDSS) assessments, as well as individual patient-level, we have been able to explore and develop more appropriate approaches. Specifically, we were able to employ more advanced statistical models, making use of all the available data and including clinically relevant patient-level characteristics as covariates in order to identify the most accurate predictive model to be applied to the RSS. Finally, we observed that to minimize 'indication bias' in relation to initiation of a DMT in the natural history cohort (British Columbia), censoring (data truncation) was more appropriate at the population (rather than individual) level.

Findings from our validation procedures indicate future feasibility with respect to obtaining reliable cost-effectiveness results in the upcoming 6 year RSS analyses. For instance, visualisation of the predicted and observed outcomes in the final model showed almost perfect overlap, with a one-covariate model, with no additional improvements from introducing further covariates. In addition, the final model was able to predict accurately the MS disease course (disability) in half of the cohort (randomly selected) having obtained the transition probabilities from the other half. These observations along with the baseline comparability of the BCMS and the RSS cohorts suggest the transition probabilities from the BCMS cohort within this model can be used to predict the untreated progression of patients in the RSS.

An additional strength of this continuous model is the ability to include all valid disability (EDSS) assessments, regardless of their exact timing, maximising data usage. We acknowledge the potential

limitations of using of a historical control from a geographically discrete population. It is possible that the natural history of MS has changed over time or that the BCMS population is not representative of a UK one. However, in British Columbia, it has been previously shown that disability progression (as measured by the EDSS) has not substantially changed overtime (1980-2009[24]). Further, we have previously shown that the use of a 'contemporary' untreated control cohort – i.e. where patients are potentially eligible for a DMT in an era when the DMTs are readily available, but remain untreated - is subject to indication bias and thus a historical control cohort, with data collected pre-DMT use, is likely to be more appropriate.[25] Although we are proposing using a dataset from Canada (as was the original RSS natural history dataset) and cannot rule out differences between the BCMS patients and the UK RSS cohort, we are reassured that the baseline features are comparable except baseline EDSS, but in the underlying Markov model we calculate the transition probabilities between EDSS 'states', and different baseline EDSS distributions would only matter if baseline EDSS as such had a prognostic value, which doesn't seem to be the case when we were looking at the rates of EDSS progression stratified by EDSS at baseline. In addition, the underlying ethnicity of the two jurisdictions was similar; around the time of the cohort selection in British Columbia, 30.2% of the population self-identified as British and within the wider BCMS database, >90% were Caucasian,[26,27] which is comparable to the UK cohort. Both cohorts may have enrolled a small number of patients with neuromyelitis optica (we estimate this to be less than 0.5% of the total[28]) because the availability of the antibody assay occurred after 2007 (and after enrolment to the RSS scheme). An additional limitation is the potential for different ways of measuring the EDSS scores between the BCMS and the UK RSS cohorts because of changes in how the EDSS is interpreted over time and also because of differences in the physicians performing the assessments.

Observational studies, such as the RSS, provide a pragmatic approach when assessing drug effectiveness in a disease such as MS. Because MS disability accrues over decades, the cost effectiveness of disease modifying treatments cannot be assessed by short-term randomised controlled trials. However, observational studies are not without their own unique challenges. Identifying and validating models to predict the untreated outcome of treated cohorts is a crucial step to measuring the long-term benefits of MS treatments. MS is the commonest cause of progressive disability in the western world, thus identification of treatments that might significantly impact long-

term disability outcomes in MS could have major cost and quality of life benefits. Additionally, any models developed here would be readily transferable to other chronic diseases.

The current model described here will form the basis for calculating the drug cost per QALY and for informing decisions on price adjustment in order to deliver the treatments cost effectively to UK MS patients in an ongoing manner. The model will be used to calculate the Hazard Ratio at which each product delivers efficacy against the NICE agreed cost per QALY and should any product fall short price reductions will be implemented by the DH.

Further work on repeated measures modelling, testing the models on other untreated appropriate MS datasets and identifying sensitivity analyses (such as the effect of drop outs, switching to a different class of DMT and the effects of treatments on backward transitions, i.e. disability improvements) are also planned.

All authors 1) made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data 2) were involved in drafting the article or revising it critically for important intellectual content and 3) were involved in the final approval of the version to be published. JP is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript. TB is the senior statistician in the scheme and the analysis of this work, was involved in the drafting and revision of this manuscript. HT, FZ and JO were responsible for the preparation of the BCMS dataset, analysis of the data, and drafting and revision of the manuscript, MB was a clinical lead for the UK risk sharing scheme, input into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript, MD is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the interpretation of the analysis and was involved in the drafting and revising of the manuscript, CD is a department of health advisor for the scheme, inputs into the scientific advisory panel, and was involved in the analysis of the data, its interpretation and was involved in the drafting and revising of the manuscript.

Acknowledgements

At the outset of the Risk-sharing Scheme agreement was reached between the UK Health Departments and the four manufacturers of the DMTs that all costs for this research project would be split 72 centres participated in the collection of data for the Risk-sharing Scheme from across the UK. Data was collected over a 10 year period and thus changes have been seen within the clinical teams. The authors are grateful to all the clinicians for the work undertaken at the centres which made the RSS possible. A list of the centres can be found at http://www.mstrust.org.uk/research/risksharingscheme/

The independent Scientific Advisory Group has been chaired by Prof Lilford and involved Prof Yoav Ben-shlomo, Prof Richard Gray and Prof Pelham Barton.

The MS Trust has performed an administrative coordinating role in the Risk-sharing Scheme and PAREXEL have performed the data collection from the sites and undertaken the analytical work.

The British Columbia MS (BCMS) database was created by Donald Paty FRCPC and Donald Studney MD FRCPC and maintained over the years by grants from various sources, including the MS Society of Canada, the MS/MRI Research Group, the US National MS Society and the Canadian Institutes of Health Research. We gratefully acknowledge the BC MS Clinic neurologists who contributed to the study through patient examination and data collection (current members listed here by primary clinic):

UBC MS Clinic: A. Traboulsee, MD, FRCPC (UBC Hospital MS Clinic Director and Head of the UBC MS Programs); A-L. Sayao, MD, FRCPC; V. Devonshire, MD, FRCPC; S. Hashimoto, MD, FRCPC (UBC and Victoria MS Clinics); J. Hooge, MD, FRCPC (UBC and Prince George MS Clinic); L. Kastrukoff, MD, FRCPC (UBC and Prince George MS Clinic); J. Oger, MD, FRCPC; Kelowna MS Clinic: D. Adams, MD, FRCPC; D. Craig, MD, FRCPC; S. Meckling, MD, FRCPC; Prince George MS Clinic: L. Daly, MD, FRCPC ; Victoria MS Clinic: O. Hrebicek, MD, FRCPC; D. Parton, MD, FRCPC; K Atwell-Pope, MD, FRCPC. The views expressed in this paper do not necessarily reflect the views of each individual acknowledged.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests:

Jacqueline Palace serves on the scientific advisory board for Charcot Foundation, and has performed advisory work for Biogen Idec, Merck Serono Ltd, Bayer Schering Pharma, Novartis Pharmaceuticals UK Ltd, Teva Pharmaceutical Industries Ltd, Gilenya, Ono Pharmaceutical Co Ltd, Primary i-research, Chugai Pharma Europe and CI Consulting. She receives research support from the MS Society, QIDIS, Merck Serono Ltd, Novartis Pharmaceuticals and Bayer Schering Pharma, plus conference expenses from Novartis and Merck Serono Ltd.

Thomas Bregenzer –as an employee of PAREXEL International (Department of Biostatistics) has been working for numerous pharmaceutical companies, including those which are participating in the UK MS Risk Sharing Scheme. no financial interests to declare.

Helen Tremlett is funded by the Multiple Sclerosis Society of Canada (Don Paty Career Development Award), Michael Smith Foundation for Health Research and is the Canada Research Chair for Neuroepidemiology and Multiple Sclerosis. She has received: research support from the US National Multiple Sclerosis Society, CIHR, and UK MS Trust; speaker honoraria and/or travel expenses to attend conferences from the Consortium of MS Centres, US National MS Society, the University of British Columbia Multiple Sclerosis Research Program, Bayer Pharmaceutical (speaker, 2010, honoraria declined), Teva Pharmaceuticals (speaker 2011), ECTRIMS (2011, 2012), UK MS Trust

(2011), and the Chesapeake Health Education Program, US Veterans Affairs (2012, honorarium declined), Novartis Canada (2012) and the US National MS Society (2012, honorarium declined). Unless otherwise stated, all speaker honoraria are either donated to an MS charity or to an unrestricted grant for use by her research group.

Joel Oger is funded by the University of British Columbia and has received grants and contracts from Biogen-Idec, Bayer, EMD Serono, Merk Serono. He has received speaker fees, Advisory Board membership and travel grants from Bayer, Novartis, Serono and the Ministry of Health of British Columbia. His lab is partially funded by the Christopher Foundation .

Feng Zhu – no financial interests to declare

Mike Boggild sits on advisory boards for Bio CSL, genzyme & Biogen Idec. Recieved sponsorship to attend international meetings from Novartis & BioCSL. Department has received funding to develop services from Biogen Idec, Genzyme and Novartis.

Martin Duddy over the past 5 years MD has received speaker honoraria, consulting fees and travel grants from, Bayer, BiogenIdec, Novartis, Merck-Serono and Teva

Charles Dobson – no financial interests to declare

Contributorship Statement

All authors 1) made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data 2) were involved in drafting the article or revising it critically for important intellectual content and 3) were involved in the final approval of the version to be published.

JP is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript. TB is the senior statistician in the scheme and the analysis of this work, was involved in the drafting and revision of this manuscript. HT, FZ and JO

were responsible for the preparation of the BCMS dataset, analysis of the data, and drafting and revision of the manuscript, MB was a clinical lead for the UK risk sharing scheme, input into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript, MD is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the interpretation of the analysis and was involved in the drafting and revising of the manuscript, CD is a department of health advisor for the scheme, inputs into the scientific advisory panel, and was involved in the analysis of the data, its interpretation and was involved in the drafting and revising of the manuscript.

Data Sharing Statement

The British Columbia MS database is held at the host institution and analysis and access to the data is limited to on site access. More detailed analysis results are available on request to the corresponding author.

References

1. National Institute for Health and Care Excellence. Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. NICE Technology Appraisal Guidance No. 32. London: NICE; 2002.

2. Department of Health. Cost effective provision of disease modifying therapies for people with multiple sclerosis. London: Health Service Circular 2002/004 London: Stationery Office; 2002.

3. Chilcott J, McCabe C, Tappenden P, O'Hagan A, Cooper NJ, Abrams K, et al. Modelling the cost-effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. BMJ 2003;326:522.

4. Hemmett L, Holmes J, Barnes M, Russell N. What drives quality of life in multiple sclerosis? QJM. 2004;97:671-6.

5. Kobelt G, Lindgren P, Parkin D, Francis DA, Johnson M, Bates D, et al. Costs and quality of life in multiple sclerosis. A cross-sectional observational study in the UK. Scandinavian Working Papers in Economics, 2000 (http://swopec.hhs.se/hastef/papers/hastef0398.pdf ; date accessed: 28-Aug-2013)

6. Boggild M, Palace J, Barton P, Ben-Shlomo Y, Bregenzer T, Dobson C, et al.. Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. BMJ 2009;339:b4677.

7. The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. Neurology. 1993;43:655-61

8. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Ann Neurol. 1996;39:285-94.

9. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. Lancet. 1998;352:1498-504.

10. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. Neurology. 1995;45:1268-76.

11. European Study Group on interferon beta-1b in secondary progressive MS. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. Lancet. 1998;352:1491-7.

12. Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-Beta-1a in MS (SPECTRIMS) Study Group. Randomized controlled trial of interferon- beta-1a in secondary progressive MS: Clinical results. Neurology. 2001;56:1496-504.

13. Tremlett H, Paty D, Devonshire V. Disability progression in MS is slower than previously reported. Neurology 2006;66:172-7.

14. Tremlett H, Zhao Y, Rieckmann P, Hutchinson M. New perspectives in the natural history of multiple sclerosis. Neurology 2010;74:2004-15.

15. Sadovnick AD, Ebers GC, Wilson RW, Paty DW. Life expectancy in patients attending multiple sclerosis clinics. Neurology. 1992;42:991-4.

16. Sweeney VP, Sadovnick AD, Brandejs V. Prevalence of multiple sclerosis in British Columbia. Can J Neurol Sci. 1986;13:47-51.

17. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol. 1983;13:227-31.

18. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Econ. 2002;11:33-42.

19. Jackson CH, Sharples LS, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. Journal of the Royal Statistical Society: Series D (The Statistician) 2003;52:193-209.

20. Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology. 2005;64:1144-51.

21. Albert P. A Markov model for sequences of ordinal data from a relapsing-remitting disease. Biometrics 1994;50:51-60.

22. Bell C, Graham J, Earnshaw S, Oleen-Burkey M, Castelli-Haley J, Johnson K., Cost-effectiveness of four immunomodulatory therapies for relapsing-remitting multiple sclerosis: a Markov model based on long-term clinical data. J Manag Care Pharm 2007;13:245-61.

23. Mandel M., Betensky R., Estimating time-to-event from longitudinal ordinal data using random-effects Markov models: application to multiple sclerosis progression. Biostatistics 2008;9:750-64.

24. Shirani A, Zhao Y, Kingwell E, Rieckmann P, Tremlett H. Temporal trends of disability progression in multiple sclerosis: findings from British Columbia, Canada (1975-2009). Mult Scler. 2012;18:442-50.

25. Shirani A, Zhao Y, Karim ME, Evans C, Kingwell E, van der Kop ML, et al. Association between use of interferon beta and progression of disability in patients with relapsing-remitting multiple sclerosis. JAMA. 2012;308:247-56.

26. http://www.bcstats.gov.bc.ca/StatisticsBySubjct/Census.aspx Specifically the 1986 Census.

27. Lourenco P, Shirani J, Saeedi J, Oger J, Schreiber W.E, the UBC MS Clinic Neurologists and Tremlett H. Oligoclonal bands and cerebrospinal fluid markers in multiple sclerosis associations with disease course and progression. Multiple Sclerosis Journal 2013 Apr;19(50):577-84.

28. http://www.msif.org/includes/documents/cm_docs/2013/m/msif-atlas-of-ms-2013-report.pdf?f=1

.

Table 1: Characteristics of patients reaching the Association of British Neurologists criteria in the British Columbia MS database after 1980 (the 'natural history' untreated comparator cohort) and the Risk-sharing Scheme cohort.

Baseline (eligible for DMT)	BCMS (1980-1995*)	RSS full cohort RSS analysis cohort**
N	898	5610 4138
Females: n (%)	666 (74.2%)	4162 (74.2%) 3125 (75.5%)
Age at baseline, years: mean (SD; range) [years]	37.2 (9.32; 18 - 69)	39.4 (9.05; 15 - 73) 38.4 (8.58; 18 - 73)
Age at onset of MS, years: mean (SD; range)	29.3 (8.65; 3 - 61)	30.5 (8.52; 5 - 68) 30.5 (8.38; 5 - 68)
Disease duration at baseline, years: mean (SD; range) [years]	7.9 (6.89; 0.2 to 38.9)	8.8 (7.47; 0 - 46) 7.7 (6.62; 0 - 41)
SPMS documented at baseline# n (%)	141 (15.7%)	772 (13.8%) -
Relapses in the last two-years prior to eligibility: median (quartiles)	2 (2 - 3)	3 (2 - 3) 3 (2 - 3)
First eligible EDSS: median (quartiles; range)	2 (1, 3.5; 0-6.5)	3.5 (2.0, 5.0; 0 - 8.0) 3.0 (2.0, 4.0; 0 - 6.5)

'Eligibility' refers to the first time a patient fulfilled the ABN criteria

**data was truncated to 1995 in the final models to minimize DMT exposure in the cohort*

*** "analysis cohort" is the subset of patients eligible for the analysis (e.g., treated patients, at least one post-baseline EDSS available etc.)*

#all were still DMT eligible

RSS=Risk-sharing Scheme ; BCMS=British Columbia MS database; SD=standard deviation; EDSS=Expanded disability status score ; DMT=disease modifying treatment

Table 2: "Goodness of fit" statistics for the ten state* disability (EDSS) Markov models

Description of each ten-state[1] disability model	Minus 2 log likelihood[2] x 1,000	Prediction errors (years 1-10)[3]		
		Cells	EDSS	Utility
No covariates	17.152	2.20	0.24	0.022
One covariate models				
Age at onset, binary	17.458	1.39	0.09	0.009
Age at onset, continuous	17.599	1.58	0.13	0.007
MSSS at baseline, binary	17.460	1.41	0.10	0.008
MSSS at baseline, continuous	17.457			
Disease duration, binary	17.462	1.33	0.10	0.009
Disease duration, continuous	17.557			
Sex	17.470	1.32	0.10	0.008
Two covariates models				
Sex and age at onset, binary	17.603	1.51	0.14	0.007
Sex and age at onset, continuous	17.618			
Age at onset and MSSS, binary	17.609	1.53	0.14	0.007
Age at onset and MSSS, continuous	17.618			
Age at onset and disease duration, binary	17.603	1.52	0.14	0.007
Age at onset and disease duration, continuous	17.618			

[1]the ten disability states refer to EDSS 0 to 9, i.e. EDSS 0 is "state 1", EDSS 1 is "state 2" etc.

[2]log likelihood statistic as calculated by 'msm' module, see Jackson[19] for details; lower values implying a better model (to be compared within each class of models, e.g. one-covariate and two-covariate models)

EDSS= Expanded disability status score ; MSSS= Multiple sclerosis status score

[3]Prediction errors, averaged over years 1-10, for (a) the EDSS distribution in individual cells, (b) average EDSS, (c) average utility (see definitions in the appendix 3, comparing the values predicted by the model with the "observed" values using the method of midpoint interpolation (see appendix 2).

Primary goodness of fit statistic is -2 log likelihood; prediction errors have only been calculated for the binary versions of the individual models except for the "final" model with age at onset as covariate where prediction errors have been calculated for both versions.

Table 3: Transition matrices for the ten state disability (EDSS) Markov model with "age at onset" as binary covariate and annual transition probabilities

age at onset < 28 yrs											
to EDSS		0	1	2	3	4	5	6	7	8	9
from EDSS	0	0.68704	0.21102	0.07195	0.02236	0.00434	0.00136	0.00176	0.00012	0.00003	0.00000
	1	0.06122	0.67867	0.16643	0.06462	0.01698	0.00474	0.00667	0.00052	0.00014	0.00001
	2	0.01692	0.12656	0.59550	0.17291	0.04537	0.01842	0.02190	0.00182	0.00054	0.00005
	3	0.00620	0.05215	0.11647	0.54386	0.09452	0.05730	0.11480	0.01070	0.00366	0.00035
	4	0.00176	0.02251	0.06617	0.12107	0.48737	0.10090	0.16644	0.02621	0.00690	0.00067
	5	0.00055	0.00562	0.02915	0.05936	0.09153	0.47268	0.28098	0.03961	0.01910	0.00143
	6	0.00012	0.00141	0.00447	0.02516	0.03208	0.04241	0.72834	0.11509	0.04566	0.00525
	7	0.00001	0.00016	0.00052	0.00260	0.00730	0.00419	0.12197	0.68145	0.16286	0.01895
	8	0.00000	0.00001	0.00004	0.00030	0.00057	0.00053	0.01884	0.05747	0.86099	0.06124
	9	0.00000	0.00000	0.00000	0.00002	0.00004	0.00004	0.00178	0.00596	0.17090	0.82125
age at onset ≥ 28 yrs											
to EDSS		0	1	2	3	4	5	6	7	8	9
from EDSS	0	0.69537	0.20294	0.07251	0.02170	0.00422	0.00137	0.00175	0.00011	0.00003	0.00000
	1	0.05826	0.69501	0.15783	0.06088	0.01638	0.00458	0.00643	0.00048	0.00013	0.00001
	2	0.01586	0.12133	0.60789	0.16796	0.04458	0.01849	0.02159	0.00174	0.00052	0.00004
	3	0.00594	0.04960	0.12006	0.54422	0.09109	0.05845	0.11649	0.01030	0.00355	0.00030
	4	0.00165	0.02214	0.06660	0.11519	0.48935	0.10388	0.16811	0.02580	0.00671	0.00056
	5	0.00052	0.00533	0.02942	0.05866	0.08736	0.48695	0.27310	0.03880	0.01883	0.00102
	6	0.00012	0.00133	0.00444	0.02497	0.03069	0.04080	0.74069	0.10897	0.04377	0.00423
	7	0.00001	0.00015	0.00052	0.00247	0.00727	0.00385	0.11684	0.69269	0.16061	0.01559
	8	0.00000	0.00001	0.00004	0.00029	0.00055	0.00050	0.01881	0.05574	0.90340	0.02066
	9	0.00000	0.00000	0.00000	0.00002	0.00004	0.00003	0.00176	0.00568	0.17414	0.81832

UK Multiple Sclerosis Risk-sharing Scheme: a new natural history dataset and an improved Markov model

The lead author* affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests:

Jacqueline Palace serves on the scientific advisory board for Charcot Foundation, and has performed advisory work for Biogen Idec, Merck Serono Ltd, Bayer Schering Pharma, Novartis Pharmaceuticals UK Ltd, Teva Pharmaceutical Industries Ltd, Gilenya, Ono Pharmaceutical Co Ltd, Primary i-research, Chugai Pharma Europe and CI Consulting. She receives research support from the MS Society, QIDIS, Merck Serono Ltd, Novartis Pharmaceuticals and Bayer Schering Pharma, plus conference expenses from Novartis and Merck Serono Ltd.

Thomas Bregenzer –as an employee of PAREXEL International (Department of Biostatistics) has been working for numerous pharmaceutical companies, including those which are participating in the UK MS Risk Sharing Scheme. no financial interests to declare.

Helen Tremlett is funded by the Multiple Sclerosis Society of Canada (Don Paty Career Development Award), Michael Smith Foundation for Health Research and is the Canada Research Chair for Neuroepidemiology and Multiple Sclerosis. She has received: research support from the US National Multiple Sclerosis Society, CIHR, and UK MS Trust; speaker honoraria and/or travel expenses to attend conferences from the Consortium of MS Centres, US National MS Society, the University of British Columbia Multiple Sclerosis Research Program, Bayer Pharmaceutical (speaker, 2010,

honoraria declined), Teva Pharmaceuticals (speaker 2011), ECTRIMS (2011, 2012), UK MS Trust (2011), and the Chesapeake Health Education Program, US Veterans Affairs (2012, honorarium declined), Novartis Canada (2012) and the US National MS Society (2012, honorarium declined). Unless otherwise stated, all speaker honoraria are either donated to an MS charity or to an unrestricted grant for use by her research group.

Joel Oger is funded by the University of British Columbia and has received grants and contracts from Biogen-Idec, Bayer, EMD Serono, Merk Serono. He has received speaker fees, Advisory Board membership and travel grants from Bayer, Novartis, Serono and the Ministry of Health of British Columbia. His lab is partially funded by the Christopher Foundation .

Feng Zhu – no financial interests to declare

Mike Boggild sits on advisory boards for Bio CSL, genzyme & Biogen Idec. Recieved sponsorship to attend international meetings from Novartis & BioCSL. Department has received funding to develop services from Biogen Idec, Genzyme and Novartis.

Martin Duddy over the past 5 years MD has received speaker honoraria, consulting fees and travel grants from, Bayer, BiogenIdec, Novartis, Merck-Serono and Teva

Charles Dobson – no financial interests to declare

Abstract

Objectives: In 2002, the UK's National Institute for Health and Care Excellence concluded that the multiple sclerosis (MS) disease modifying therapies; interferon-β and glatiramer acetate, were not cost-effective over the short term but recognised that reducing disability over the longer-term might dramatically improve the cost -effectiveness. The UK Risk-sharing Scheme (RSS) was established to ensure cost-effective provision by prospectively collecting disability-related data from UK treated MS patients and comparing findings to a natural history (untreated) cohort. However, deficiencies were found in the originally selected untreated cohort and the resulting analytical approach. This study aims to identify a more suitable natural history comparator cohort and to develop a robust analytical approach using the new cohort

Design: The Scientific Advisory Group review, recommended the British Columbia Multiple Sclerosis (BCMS) database, Canada, as providing a more suitable natural history comparator cohort. Transition probabilities were derived and different Markov models (discrete and continuous) with and without baseline covariates were applied.

Setting: MS clinics and analysis groups in Canada and the UK.

Participants: From the BCMS database, 898 'untreated' MS patients considered eligible for drug treatment based on the UK's Association of British Neurologists criteria.

Outcome measure: The predicted disability, as measured by the Expanded Disability Status Scale (EDSS) score was collected and assessed for goodness of fit when compared to actual outcome.

Results: The BCMS untreated cohort contributed 7335 EDSS scores over a median 6.4 years (6357 "transitions" where EDSS values were recorded at consecutive visits) during the study period (1980-1995). A continuous Markov model with "age at onset" as a binary covariate was deemed the most suitable model for future RSS analysis.

Conclusion: A new untreated MS cohort from British Columbia, Canada has been selected and will be modelled using a continuous Markov model with age as a baseline covariate. This approach will now be applied to the treated UK RSS MS cohort for future price adjustment calculations.

Strengths and limitations of this study

Strengths of this study:

- Identification of a new natural history cohort for the UKRSS, consisting of untreated MS patients in an era when disease-modifying drugs for MS were not available, minimizing potential selection bias
- Identification and validation of a Markov model for disease progression in MS which can be applied to data collected in clinical practice over multiple years of follow up.
- The identification of an analytical model which can use data collected at any time point within the follow up period.

Limitations of this study:

- The study related to observational data collected in clinical practice; unseen or unmeasured confounding cannot be adjusted for.
- Different techniques to assess effectiveness of drugs in observational studies such as matching on propensity scores cannot be directly compared to this methodology.

Introduction

In January 2002, the UK's National Institute of Health and Care Excellence (NICE) opted not to recommend the use of the disease modifying therapies (DMTs) interferon-β and glatiramer acetate for multiple sclerosis (MS) on the basis of cost-effectiveness analyses using data derived from the pivotal 2-3 year randomised controlled trials.[1] However, they recognised that uncertainties over the assumptions made in the modelling could unpredictably influence the long-term estimates of cost effectiveness. Thus in February 2002 the UK's Department of Health launched the 'Risk-sharing Scheme' (RSS)[2] with a circular entitled the "Cost effective provision of disease modifying therapies for people with multiple sclerosis" in collaboration with the Association of British Neurologists (ABN), the MS Trust, the MS Society and the pharmaceutical companies manufacturing interferon-β and glatiramer acetate. Between 2002 and 2005 the scheme enrolled over 5,000 MS patients initiating a DMT in the UK, with the aim of measuring their disability annually over a ten year period.

The principle In the RSS is to use a Markov model to predict, for each DMT separately, the expected movement of patients between the EDSS states both "on" and "off" treatment. For patients "off" treatment, the model uses a matrix of transition probabilities derived from the actual progressions seen in the 'natural history' comparator cohort. These transition matrices are modified for patients "on" treatment by multiplying by the hazard ratio (relative rate of disease progression) derived separately for each DMT from the pivotal randomized controlled clinical trials. The model then predicts how the distribution of patients will evolve over a 20-year horizon, starting with the actual distribution at baseline for the primary analysis RSS cohort. Comparing the average observed loss of utility (average utility-weighted disease progression) for patients in the RSS to the expected loss calculated by the Markov model for patients "on" treatment; it is calculated as follows. The *expected* 'benefit' of treatment (with a specific DMT) is the "hypothetical" difference between the expected outcome without treatment and with treatment, as calculated in each case from the Markov model. The *actual* 'benefit' of treatment is the "observed" difference between the expected outcome without treatment and the actual outcome with treatment. The 'deviation' of the actual benefit from the expected is the primary outcome measure and calculated as a percentage of the expected benefit. This measure can have negative or positive values so that a negative deviation implies that the observed benefit was greater than predicted, a positive deviation suggesting that it was worse than

predicted and a value of 0 indicating that it was exactly as predicted. A positive deviation beyond the level agreed (confidential and individual between each pharmaceutical company and the Department of Health) would lead to a price adjustment down to achieve the target cost effectiveness. Details can be found in the Health Service Circular[2].

The original cost effectiveness model[3] produced a target outcome based upon transition probabilities obtained from a pre-existing natural history (DMT naive) cohort of patients from London, Ontario, Canada along with hazard ratios from the pivotal randomised control trials (unpublished data provided to the Department of Health by the manufacturers). Complementary quality of life data collected by the MS Trust[4] and cost data from Kobelt et al[5] were used to populate the cost-effectiveness model. The targets ensured that the UK's National Health Service benchmark of £36,000 (46,000 Euro / 56,000 US dollars) per quality adjusted life year (QALY) was reached over a 20 year projection, based on a planned 10 year follow up period within the RSS with 2 yearly interim analyses. Before being allowed to enter the scheme, the costs of each drug was assessed against the NICE bench mark over a 20 year time horizon. Price reductions were implemented to ensure each product reached the target cost per QALY using the original NICE calculations[3], an average 13.7% price reduction was achieved for the NHS at the outset of the Scheme.

The two year analysis revealed significant inconsistencies in a number of sensitivity analyses.[6] Depending on the underlying assumptions, some analyses suggested that observed disability progression in the treated cohort was worse than that predicted from the historical untreated cohort while others demonstrated the contrary effect. A detrimental effect of DMT did not match the described effect on short term, 2-3 year, disability seen in the randomised placebo controlled trials.[7-12] With the predetermined analytical approach (based on a discrete Markov model) appearing to produce unreliable results with wide variation, a decision was made to postpone any decision on cost effectiveness to allow for a reappraisal of the process and to reconsider whether the statistical models and control data chosen were "fit for purpose."

In retrospect, both the control data set and analysis model selected, when setting up the RSS, were found to have intrinsic flaws that made them unsuitable for the task.[6] The natural history cohort (from London, Ontario, Canada) was unexpectedly found to contain retrospectively smoothed disability data (rather than actual, real-time collected disability scores), censoring any improvement in EDSS.

Comparing our uncensored treated cohort to data retrospectively smoothed in this way would have the effect of unpredictably underestimating any treatment effect. In addition, individual-level patient data were not available from the London, Ontario cohort, which prevented precise baseline matching between the two cohorts, limiting our validation of the underlying (Markov) model for disease progression. Furthermore, there were only 342 patients matching the ABN prescribing criteria from which to generate the models.

This paper outlines the development of a more appropriate analysis plan and the choice of a cohort fit for the needs of the scheme. The method described will be applied in the 4 and 6 year cost-effectiveness analyses. The analysis plan was approved by the scheme's independent Scientific Advisory Group in December 2012 in advance of unlocking the newly collected 4 and 6 year UK Risk-sharing Scheme data planned for autumn 2013.

Methods

Identification of a new multiple sclerosis natural history dataset

The Scientific Advisory Group undertook a detailed examination of all the available dataset through literature reviews, expert opinion, discussion with the clinical leads for the RSS and discussion with the Sylvia Lawry Centre for Multiple Sclerosis Research, Germany (http://www.slcmsr.net). Selection criteria included availability of Expanded Disability Status Scale (EDSS) score measurements and access to the unprocessed (actual) scores (i.e. no data smoothing or other data manipulation). Other factors considered were size of the database, prospective data collection and length of follow-up, and the broader setting such as a close match to the UK in terms of the health system and MS prevalence in the underlying population. Whilst no single perfect dataset existed the British Columbia Multiple Sclerosis (BCMS) database, Canada (est. 1980) was identified as the best natural history comparator cohort for our purposes.[13,14] In this dataset – as in the RSS – actual EDSS scores were recorded prospectively. It is estimated to capture 80% of the BC MS population[15,16] and as such is considered representative of the wider MS population. EDSS scores were recorded by MS specialist neurologists after a face-to-face consultation with the patient; this typically occurred at the annual MS clinic visit. Patient data was not truncated if secondary-progressive MS was reached; i.e. all relapse-onset MS patients and their respective EDSS scores were considered eligible. By 2004, the database had

records for over 5900 patients spanning 28 years (>25,000 cumulative years) of prospective follow-up. Until 1996 DMTs were not widely available in British Columbia.

Patient and data selection from the BCMS database.

In order to generate relevant data for our needs, patients were only selected from the BCMS database if they fulfilled the 2001 Association of British Neurologists (ABN) criteria for interferon-β and glatiramer acetate (IFN- β/GA) use (adapted from Appendix IV Health Service Circular 2002/004), defined as: EDSS≤6.5; ≥18 years old; two relapses in the last 2 calendar years.

Baseline for each patient was the 'first eligibility date,' meaning the first date at which a patient fulfilled the ABN eligibility criteria. Only patients with definite MS (Poser criteria[17]) and a minimum of two EDSS scores at least 9 months apart were considered.

In order to be comparable with the RSS data the following adjustments and selection were applied:

1) EDSS scores taken during a relapse or when disability was affected by other factors considered largely unrelated to MS (e.g. hip fracture) were excluded.

2) For the original discrete Markov model (see below) as well as visual presentation of the yearly descriptive data (see under results), annual EDSS scores were needed. However, as is typical in clinical practice, not all visits / EDSS assessments occurred at exactly yearly intervals and the exclusion of some EDSS scores (e.g. due to a relapse or hip fracture) also affected the availability of a yearly score. Therefore, data was selected such that only EDSS scores one year apart (+/- three calendar months) were considered. See appendix 1 for further details.

3) For the continuous Markov model, (see below) all eligible EDSS scores were used regardless of their measurement interval i.e. no yearly data selection, as in (2), was needed.

4) All patient data was truncated to the end of 1995 (i.e. the last full year in which the DMTs were not widely available in BC Although initially it was planned to truncate individual patient profiles only once a DMT was initiated (in order to maximise the number of EDSS assessments), even when this extended past 1995 when treatment would have been available). It became apparent that this

introduced a bias into the data, likely related to 'indication bias,' whereby patients 'doing well' would be less likely to start a DMT.

Analysis

The primary purpose of the analytical work was to find the best statistical model able to predict EDSS progression in a natural history cohort based on entry demographic and clinical data. The following models were applied in the current study and their performances were critically evaluated.

a) The discrete Markov model[18] as in the original 2 year analysis[6] i.e. disability scores (EDSS) had to be measured at discrete, fixed time points.

b) A continuous Markov model allowing for EDSS scores to be collected at any time, i.e. at any unevenly spaced time intervals, as is typical in clinical practice.[19] Such a model also allowed covariates to be included. This model allows for more complete use of EDSS scores collected at irregular time intervals both in the BCMS and RSS cohorts.

With regard to the "MS course" (i.e. relapsing remitting vs. secondary progressive) as a potential covariate, we did not distinguish between these disease states when developing the Markov models because secondary progressive MS is simply a later stage of the relapsing remitting form of the disease and the transition has considerable overlap.

Predicting outcomes in the continuous Markov model (b)

A multi-state model algorithm ('R' library 'msm'[19]) allows the EDSS distribution to be predicted at any time *t*. See appendix 2 for further details).

To keep computations feasible, only integer EDSS values were used and fractional values rounded down (i.e. EDSS 1.5 was scored as 1, 2.5 was scored as 2 etc.). Moreover, 'msm' as a tool for *multi-state modelling* requires a consecutive numbering of (disease) states, starting with "1". Therefore the (rounded down) EDSS 0 became 'state 1', EDSS 1 'state 2' etc., leading to the ten EDSS 'states' (1–10) representing EDSS 0 –9. Transition probability and intensity matrices as the output of these models were then used to predict disease progression in terms of EDSS as follows.

Covariates considered in the models

The selection of potential covariates by the scientific advisory group was based upon; (a) those which were reliably and consistently recorded in both the BCMS and the RSS database and (b) *a priori* knowledge of those associated with the outcome of disability progression. As a result, sex, age at MS symptom onset, as well as disease duration and disability (EDSS) – combined into a Multiple Sclerosis Severity Score (MSSS)[20] were tested in the continuous Markov model with up to two covariates. In addition, for the more promising models an alternative model was considered with dichotomous covariates (split at the median) replacing the continuous variables. This has the advantage that the resulting model can be formulated as the aggregate of a small number of discrete Markov models, so computations can be carried out without requiring special software, especially since the EDSS values in the RSS have been collected at strict yearly intervals, as opposed to the BCMS data which was based on routine clinical practice, and therefore do not necessitate a continuous model.

Critical evaluation of the models was performed using the following validation techniques, with the goal being to identify the most appropriate model to represent the natural progression of MS. See appendix 3 for further details.

1) Transition probabilities derived from the complete eligible, BCMS natural history data were applied to the baseline data to predict outcomes over the subsequent 10 years to assess how well it matched the observed data from which the model was derived.

2) The BCMS dataset was repeatedly randomly divided into two subsets of equal size, with one half only being used to derive transition probabilities (as in #1). The probabilities derived from this half were then applied to the baseline characteristics of the second half, generating a model whose goodness of fit could be judged against the actual ,observed 10 year disability data of this second half.

Measuring Goodness of Fit

Goodness of fit was assessed via visual inspection of the graphical displays as well as numerically. These included progression over time (mean EDSS profiles) for the cohort as a whole as well as comparisons with the proportions in a particular EDSS state over time.

For the numerical assessment a classical mean square prediction error (weighted root mean square over years of the prediction error in the average quantity shown, weighted by the number of patients contributing data in the given year) and the likelihood, resulting from the maximum likelihood algorithm were calculated for each of the covariate models to allow comparison.

This study was approved by the University of British Columbia's Clinical Research Ethics Board (H08-01544)'

Results

Data Description

The baseline demographics showed the BCMS and RSS cohorts to be remarkably well matched. Patient characteristics are shown in table 1.

The natural history BCMS comparator dataset comprised of 898 patient profiles with 7335 EDSS scores providing 6357 transitions between consecutive EDSS states, i.e 6357 'events' where EDSS values were recorded at consecutive visits. In any given "transition," a patient's EDSS could increase, decrease or stay the same

Discrete Markov model

When applying the discrete Markov model to the BCMS reference data, the goodness of fit was unsatisfactory, underestimating EDSS in earlier years and overestimating in later years (see figure 1). Consequently, the discrete Markov model was no longer considered appropriate, and development of a continuous Markov model was pursued.

Continuous Markov models

The following continuous 10 state Markov models (EDSS 0 to 9), with and without covariates, were evaluated:

1. Model without covariates

2. One covariate model with age at onset*
3. One covariate model with MSSS* at baseline
4. One covariate model with disease duration* at baseline
5. One covariate model with sex
6. Two covariate model: sex and age at onset*
7. Two covariate model: MSSS* at baseline and age at onset*
8. Two covariate model: disease duration* at baseline and age at onset*

- *two variants were implemented: continuous (original) data and a 'binary' version with the median used for categorisation.

There was a systematic deviation with a slight overestimation when the continuous Markov model without covariates was applied (figure 2). Hence this model was not considered further.

After validation was repeated for all covariate models (table 2), it was noted that inclusion of a second covariate did not reveal any additional benefits. With one covariate, the model with "age at onset" as a binary covariate (median: 28 years) was selected because it displayed the smallest -2 log likelihood and minimal EDSS prediction error, see table 2. Further, the goodness of fit was acceptable when comparing the predicted and observed EDSS profiles, as shown in Figure 3a. A more detailed comparison of observed and expected proportions 'per EDSS state' is shown in Figure 3b which confirmed that no systematic deviations were present which might otherwise have been cancelled out when only looking at an average EDSS profile. It was concluded that only random fluctuation remained, and a systematic deviation was no longer visible. When comparing figure 1 with figures 2 or 3 it should be noted that the former is based on the *annual* EDSS data which were obtained as described in Appendix 1 while figures 2 and 3 show the EDSS at any time *t*, i.e. not necessarily when an observation was recorded (while the continuous Markov model takes into account all observations at any time *t* it is not straightforward to define what the 'observed EDSS' at any time *t* is in a graphic representation; for details on how to define and calculate what is the observed EDSS at a given time see Appendix 2).

Using this 'best' model, transition probabilities were extracted from half of the BCMS cohort and applied to the other half. This gave good predictions, with the mean EDSS profiles (observed versus predicted) being similar to each other and to those of the entire cohort.

In summary, the continuous Markov model with a single covariate - onset age - was considered the model of choice to be used in future RSS analysis. The respective transition matrices are presented in Table 3.

Discussion

This paper outlines the successful identification of a more suitable natural history cohort for the UK MS risk-sharing scheme, with the British Columbia, Canada dataset now replacing the London, Ontario, Canada cohort in the RSS analysis plan. The analytical work is based on a Markov model which has been frequently used for ordinal data from relapsing (remitting) diseases, especially MS.[21,22,23]

Further, because use of the British Columbian data has now allowed access to a richer dataset, including full access to original, 'real-time' disability (EDSS) assessments, as well as individual patient-level, we have been able to explore and develop more appropriate approaches. Specifically, we were able to employ more advanced statistical models, making use of all the available data and including clinically relevant patient-level characteristics as covariates in order to identify the most accurate predictive model to be applied to the RSS. Finally, we observed that to minimize 'indication bias' in relation to initiation of a DMT in the natural history cohort (British Columbia), censoring (data truncation) was more appropriate at the population (rather than individual) level.

Findings from our validation procedures indicate future feasibility with respect to obtaining reliable cost-effectiveness results in the upcoming 6 year RSS analyses. For instance, visualisation of the predicted and observed outcomes in the final model showed almost perfect overlap, with a one-covariate model, with no additional improvements from introducing further covariates. In addition, the final model was able to predict accurately the MS disease course (disability) in half of the cohort (randomly selected) having obtained the transition probabilities from the other half. These observations along with the baseline comparability of the BCMS and the RSS cohorts suggest the transition probabilities from the BCMS cohort within this model can be used to predict the untreated progression of patients in the RSS.

An additional strength of this continuous model is the ability to include all valid disability (EDSS) assessments, regardless of their exact timing, maximising data usage. We acknowledge the potential

limitations of using of a historical control from a geographically discrete population. It is possible that the natural history of MS has changed over time or that the BCMS population is not representative of a UK one. However, in British Columbia, it has been previously shown that disability progression (as measured by the EDSS) has not substantially changed overtime (1980-2009[24]). Further, we have previously shown that the use of a 'contemporary' untreated control cohort – i.e. where patients are potentially eligible for a DMT in an era when the DMTs are readily available, but remain untreated - is subject to indication bias and thus a historical control cohort, with data collected pre-DMT use, is likely to be more appropriate.[25] Although we are proposing using a dataset from Canada (as was the original RSS natural history dataset) and cannot rule out differences between the BCMS patients and the UK RSS cohort, we are reassured that the baseline features are comparable except baseline EDSS, but in the underlying Markov model we calculate the transition probabilities between EDSS 'states', and different baseline EDSS distributions would only matter if baseline EDSS as such had a prognostic value, which doesn't seem to be the case when we were looking at the rates of EDSS progression stratified by EDSS at baseline. In addition, the underlying ethnicity of the two jurisdictions was similar; around the time of the cohort selection in British Columbia, 30.2% of the population self-identified as British and within the wider BCMS database, >90% were Caucasian,[26,27] which is comparable to the UK cohort. Both cohorts may have enrolled a small number of patients with neuromyelitis optica (we estimate this to be less than 0.5% of the total[28]) because the availability of the antibody assay occurred after 2007 (and after enrolment to the RSS scheme). An additional limitation is the potential for different ways of measuring the EDSS scores between the BCMS and the UK RSS cohorts because of changes in how the EDSS is interpreted over time and also because of differences in the physicians performing the assessments.

Observational studies, such as the RSS, provide a pragmatic approach when assessing drug effectiveness in a disease such as MS. Because MS disability accrues over decades, the cost effectiveness of disease modifying treatments cannot be assessed by short-term randomised controlled trials. However, observational studies are not without their own unique challenges. Identifying and validating models to predict the untreated outcome of treated cohorts is a crucial step to measuring the long-term benefits of MS treatments. MS is the commonest cause of progressive disability in the western world, thus identification of treatments that might significantly impact long-

term disability outcomes in MS could have major cost and quality of life benefits. Additionally, any models developed here would be readily transferable to other chronic diseases.

The current model described here will form the basis for calculating the drug cost per QALY and for informing decisions on price adjustment in order to deliver the treatments cost effectively to UK MS patients in an ongoing manner. The model will be used to calculate the Hazard Ratio at which each product delivers efficacy against the NICE agreed cost per QALY and should any product fall short price reductions will be implemented by the DH.

Further work on repeated measures modelling, testing the models on other untreated appropriate MS datasets and identifying sensitivity analyses (such as the effect of drop outs, switching to a different class of DMT and the effects of treatments on backward transitions, i.e. disability improvements) are also planned.

All authors 1) made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data 2) were involved in drafting the article or revising it critically for important intellectual content and 3) were involved in the final approval of the version to be published. JP is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript. TB is the senior statistician in the scheme and the analysis of this work, was involved in the drafting and revision of this manuscript. HT, FZ and JO were responsible for the preparation of the BCMS dataset, analysis of the data, and drafting and revision of the manuscript, MB was a clinical lead for the UK risk sharing scheme, input into the scientific advisory panel, was involved in the identification of the alternative dataset, interpretation of the analysis and was involved in the drafting and revising of the manuscript, MD is a clinical lead for the UK risk sharing scheme, inputs into the scientific advisory panel, was involved in the interpretation of the analysis and was involved in the drafting and revising of the manuscript, CD is a department of health advisor for the scheme, inputs into the scientific advisory panel, and was involved in the analysis of the data, its interpretation and was involved in the drafting and revising of the manuscript.

Acknowledgements

At the outset of the Risk-sharing Scheme agreement was reached between the UK Health Departments and the four manufacturers of the DMTs that all costs for this research project would be split 72 centres participated in the collection of data for the Risk-sharing Scheme from across the UK. Data was collected over a 10 year period and thus changes have been seen within the clinical teams. The authors are grateful to all the clinicians for the work undertaken at the centres which made the RSS possible. A list of the centres can be found at http://www.mstrust.org.uk/research/risksharingscheme/

The independent Scientific Advisory Group has been chaired by Prof Lilford and involved Prof Yoav Ben-shlomo, Prof Richard Gray and Prof Pelham Barton.

The MS Trust has performed an administrative coordinating role in the Risk-sharing Scheme and PAREXEL have performed the data collection from the sites and undertaken the analytical work.

The British Columbia MS (BCMS) database was created by Donald Paty FRCPC and Donald Studney MD FRCPC and maintained over the years by grants from various sources, including the MS Society of Canada, the MS/MRI Research Group, the US National MS Society and the Canadian Institutes of Health Research. We gratefully acknowledge the BC MS Clinic neurologists who contributed to the study through patient examination and data collection (current members listed here by primary clinic):

UBC MS Clinic: A. Traboulsee, MD, FRCPC (UBC Hospital MS Clinic Director and Head of the UBC MS Programs); A-L. Sayao, MD, FRCPC; V. Devonshire, MD, FRCPC; S. Hashimoto, MD, FRCPC (UBC and Victoria MS Clinics); J. Hooge, MD, FRCPC (UBC and Prince George MS Clinic); L. Kastrukoff, MD, FRCPC (UBC and Prince George MS Clinic); J. Oger, MD, FRCPC; Kelowna MS Clinic: D. Adams, MD, FRCPC; D. Craig, MD, FRCPC; S. Meckling, MD, FRCPC; Prince George MS Clinic: L. Daly, MD, FRCPC ; Victoria MS Clinic: O. Hrebicek, MD, FRCPC; D. Parton, MD, FRCPC; K Atwell-Pope, MD, FRCPC. The views expressed in this paper do not necessarily reflect the views of each individual acknowledged.

References

1. National Institute for Health and Care Excellence. Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. NICE Technology Appraisal Guidance No. 32. London: NICE; 2002.

2. Department of Health. Cost effective provision of disease modifying therapies for people with multiple sclerosis. London: Health Service Circular 2002/004 London: Stationery Office; 2002.

3. Chilcott J, McCabe C, Tappenden P, O'Hagan A, Cooper NJ, Abrams K, et al. Modelling the cost-effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. BMJ 2003;326:522.

4. Hemmett L, Holmes J, Barnes M, Russell N. What drives quality of life in multiple sclerosis? QJM. 2004;97:671-6.

5. Kobelt G, Lindgren P, Parkin D, Francis DA, Johnson M, Bates D, et al. Costs and quality of life in multiple sclerosis. A cross-sectional observational study in the UK. Scandinavian Working Papers in Economics, 2000 (http://swopec.hhs.se/hastef/papers/hastef0398.pdf ; date accessed: 28-Aug-2013)

6. Boggild M, Palace J, Barton P, Ben-Shlomo Y, Bregenzer T, Dobson C, et al.. Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. BMJ 2009;339:b4677.

7. The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. Neurology. 1993;43:655-61

8. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Ann Neurol. 1996;39:285-94.

9. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. Lancet. 1998;352:1498-504.

10. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. Neurology. 1995;45:1268-76.

11. European Study Group on interferon beta-1b in secondary progressive MS. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. Lancet. 1998;352:1491-7.

12. Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-Beta-1a in MS (SPECTRIMS) Study Group. Randomized controlled trial of interferon- beta-1a in secondary progressive MS: Clinical results. Neurology. 2001;56:1496-504.

13. Tremlett H, Paty D, Devonshire V. Disability progression in MS is slower than previously reported. Neurology 2006;66:172-7.

14. Tremlett H, Zhao Y, Rieckmann P, Hutchinson M. New perspectives in the natural history of multiple sclerosis. Neurology 2010;74:2004-15.

15. Sadovnick AD, Ebers GC, Wilson RW, Paty DW. Life expectancy in patients attending multiple sclerosis clinics. Neurology. 1992;42:991-4.

16. Sweeney VP, Sadovnick AD, Brandejs V. Prevalence of multiple sclerosis in British Columbia. Can J Neurol Sci. 1986;13:47-51.

17. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol. 1983;13:227-31.

18. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Econ. 2002;11:33-42.

19. Jackson CH, Sharples LS, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. Journal of the Royal Statistical Society: Series D (The Statistician) 2003;52:193-209.

20. Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology. 2005;64:1144-51.

21. Albert P. A Markov model for sequences of ordinal data from a relapsing-remitting disease. Biometrics 1994;50:51-60.

22. Bell C, Graham J, Earnshaw S, Oleen-Burkey M, Castelli-Haley J, Johnson K., Cost-effectiveness of four immunomodulatory therapies for relapsing-remitting multiple sclerosis: a Markov model based on long-term clinical data. J Manag Care Pharm 2007;13:245-61.

23. Mandel M., Betensky R., Estimating time-to-event from longitudinal ordinal data using random-effects Markov models: application to multiple sclerosis progression. Biostatistics 2008;9:750-64.

24. Shirani A, Zhao Y, Kingwell E, Rieckmann P, Tremlett H. Temporal trends of disability progression in multiple sclerosis: findings from British Columbia, Canada (1975-2009). Mult Scler. 2012;18:442-50.

25. Shirani A, Zhao Y, Karim ME, Evans C, Kingwell E, van der Kop ML, et al. Association between use of interferon beta and progression of disability in patients with relapsing-remitting multiple sclerosis. JAMA. 2012;308:247-56.

26. http://www.bcstats.gov.bc.ca/StatisticsBySubjct/Census.aspx Specifically the 1986 Census.

27. Lourenco P, Shirani J, Saeedi J, Oger J, Schreiber W.E, the UBC MS Clinic Neurologists and Tremlett H. Oligoclonal bands and cerebrospinal fluid markers in multiple sclerosis associations with disease course and progression. Multiple Sclerosis Journal 2013 Apr;19(50):577-84.

28. http://www.msif.org/includes/documents/cm_docs/2013/m/msif-atlas-of-ms-2013-report.pdf?f=1

.

Table 1: Characteristics of patients reaching the Association of British Neurologists criteria in the British Columbia MS database after 1980 (the 'natural history' untreated comparator cohort) and the Risk-sharing Scheme cohort.

Baseline (eligible for DMT)	BCMS (1980-1995*)	RSS full cohort RSS analysis cohort**
N	898	5610 4138
Females: n (%)	666 (74.2%)	4162 (74.2%) 3125 (75.5%)
Age at baseline, years: mean (SD; range) [years]	37.2 (9.32; 18 - 69)	39.4 (9.05; 15 - 73) 38.4 (8.58; 18 - 73)
Age at onset of MS, years: mean (SD; range)	29.3 (8.65; 3 - 61)	30.5 (8.52; 5 - 68) 30.5 (8.38; 5 - 68)
Disease duration at baseline, years: mean (SD; range) [years]	7.9 (6.89; 0.2 to 38.9)	8.8 (7.47; 0 - 46) 7.7 (6.62; 0 - 41)
SPMS documented at baseline# n (%)	141 (15.7%)	772 (13.8%) -
Relapses in the last two-years prior to eligibility: median (quartiles)	2 (2 - 3)	3 (2 - 3) 3 (2 - 3)
First eligible EDSS: median (quartiles; range)	2 (1, 3.5; 0-6.5)	3.5 (2.0, 5.0; 0 - 8.0) 3.0 (2.0, 4.0; 0 - 6.5)

'Eligibility' refers to the first time a patient fulfilled the ABN criteria

**data was truncated to 1995 in the final models to minimize DMT exposure in the cohort*

*** "analysis cohort" is the subset of patients eligible for the analysis (e.g., treated patients, at least one post-baseline EDSS available etc.)*

#all were still DMT eligible

RSS=Risk-sharing Scheme ; BCMS=British Columbia MS database; SD=standard deviation;

EDSS=Expanded disability status score ; DMT=disease modifying treatment

Table 2: "Goodness of fit" statistics for the ten state* disability (EDSS) Markov models

Description of each ten-state[1] disability model	Minus 2 log likelihood[2] x 1,000	Prediction errors 1-10)[3] Cells	EDSS	(years Utility
No covariates	17.152	2.20	0.24	0.022
One covariate models				
Age at onset, binary	17.458	1.39	0.09	0.009
Age at onset, continuous	17.599	1.58	0.13	0.007
MSSS at baseline, binary	17.460	1.41	0.10	0.008
MSSS at baseline, continuous	17.457			
Disease duration, binary	17.462	1.33	0.10	0.009
Disease duration, continuous	17.557			
Sex	17.470	1.32	0.10	0.008
Two covariates models				
Sex and age at onset, binary	17.603	1.51	0.14	0.007
Sex and age at onset, continuous	17.618			
Age at onset and MSSS, binary	17.609	1.53	0.14	0.007
Age at onset and MSSS, continuous	17.618			
Age at onset and disease duration, binary	17.603	1.52	0.14	0.007
Age at onset and disease duration, continuous	17.618			

[1]the ten disability states refer to EDSS 0 to 9, i.e. EDSS 0 is "state 1", EDSS 1 is "state 2" etc.

[2]log likelihood statistic as calculated by 'msm' module, see Jackson[19] for details; lower values implying a better model (to be compared within each class of models, e.g. one-covariate and two-covariate models)

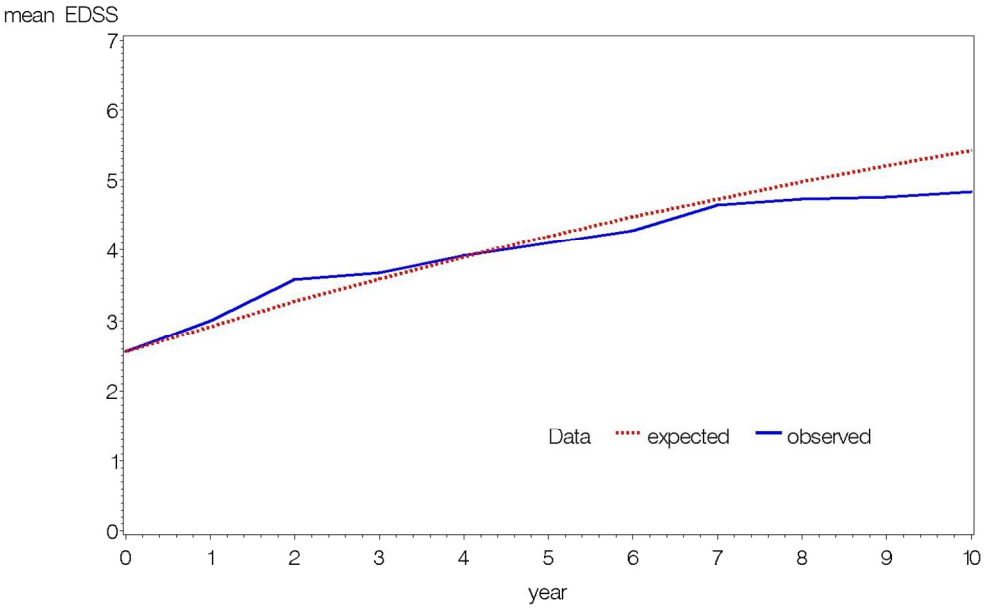
EDSS= Expanded disability status score ; MSSS= Multiple sclerosis status score

[3]Prediction errors, averaged over years 1-10, for (a) the EDSS distribution in individual cells, (b) average EDSS, (c) average utility (see definitions in the appendix 3, comparing the values predicted by the model with the "observed" values using the method of midpoint interpolation (see appendix 2).

Primary goodness of fit statistic is -2 log likelihood; prediction errors have only been calculated for the binary versions of the individual models except for the "final" model with age at onset as covariate where prediction errors have been calculated for both versions.

Table 3: Transition matrices for the ten state disability (EDSS) Markov model with "age at onset" as binary covariate and annual transition probabilities

age at onset < 28 yrs											
to EDSS		0	1	2	3	4	5	6	7	8	9
from EDSS	0	0.68704	0.21102	0.07195	0.02236	0.00434	0.00136	0.00176	0.00012	0.00003	0.00000
	1	0.06122	0.67867	0.16643	0.06462	0.01698	0.00474	0.00667	0.00052	0.00014	0.00001
	2	0.01692	0.12656	0.59550	0.17291	0.04537	0.01842	0.02190	0.00182	0.00054	0.00005
	3	0.00620	0.05215	0.11647	0.54386	0.09452	0.05730	0.11480	0.01070	0.00366	0.00035
	4	0.00176	0.02251	0.06617	0.12107	0.48737	0.10090	0.16644	0.02621	0.00690	0.00067
	5	0.00055	0.00562	0.02915	0.05936	0.09153	0.47268	0.28098	0.03961	0.01910	0.00143
	6	0.00012	0.00141	0.00447	0.02516	0.03208	0.04241	0.72834	0.11509	0.04566	0.00525
	7	0.00001	0.00016	0.00052	0.00260	0.00730	0.00419	0.12197	0.68145	0.16286	0.01895
	8	0.00000	0.00001	0.00004	0.00030	0.00057	0.00053	0.01884	0.05747	0.86099	0.06124
	9	0.00000	0.00000	0.00000	0.00002	0.00004	0.00004	0.00178	0.00596	0.17090	0.82125
age at onset ≥ 28 yrs											
to EDSS		0	1	2	3	4	5	6	7	8	9
from EDSS	0	0.69537	0.20294	0.07251	0.02170	0.00422	0.00137	0.00175	0.00011	0.00003	0.00000
	1	0.05826	0.69501	0.15783	0.06088	0.01638	0.00458	0.00643	0.00048	0.00013	0.00001
	2	0.01586	0.12133	0.60789	0.16796	0.04458	0.01849	0.02159	0.00174	0.00052	0.00004
	3	0.00594	0.04960	0.12006	0.54422	0.09109	0.05845	0.11649	0.01030	0.00355	0.00030
	4	0.00165	0.02214	0.06660	0.11519	0.48935	0.10388	0.16811	0.02580	0.00671	0.00056
	5	0.00052	0.00533	0.02942	0.05866	0.08736	0.48695	0.27310	0.03880	0.01883	0.00102
	6	0.00012	0.00133	0.00444	0.02497	0.03069	0.04080	0.74069	0.10897	0.04377	0.00423
	7	0.00001	0.00015	0.00052	0.00247	0.00727	0.00385	0.11684	0.69269	0.16061	0.01559
	8	0.00000	0.00001	0.00004	0.00029	0.00055	0.00050	0.01881	0.05574	0.90340	0.02066
	9	0.00000	0.00000	0.00000	0.00002	0.00004	0.00003	0.00176	0.00568	0.17414	0.81832



Transition probabilities obtained from the BCC dataset using the discrete Markov model were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted mean EDSS outcome (red) and compared to the observed mean EDSS course of the cohort (blue).
402x264mm (96 x 96 DPI)

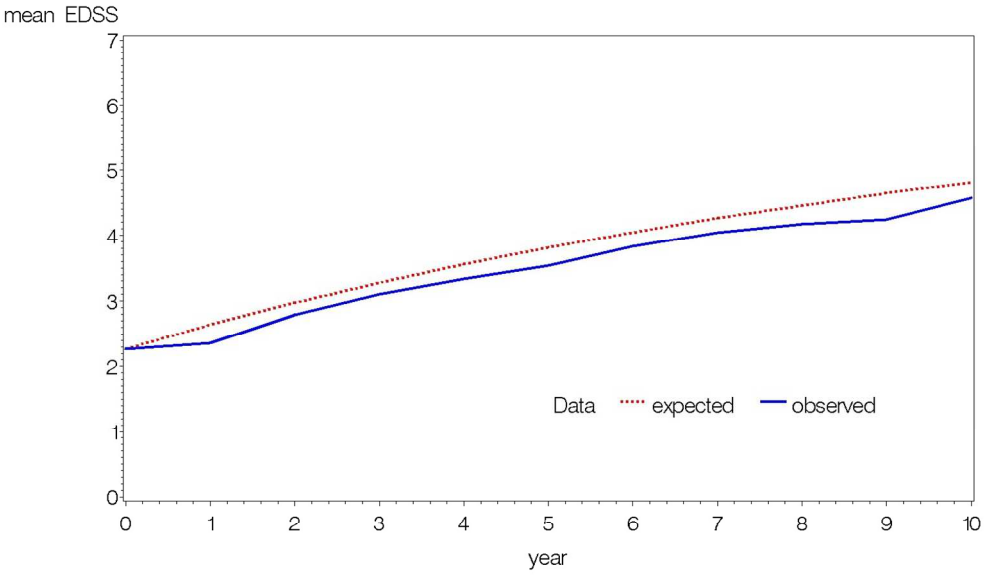


Figure 2: Transition probabilities obtained from the BCC dataset using the continuous Markov model were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted mean EDSS outcome (red) and compared to the observed mean EDSS course of the cohort (blue).
402x264mm (96 x 96 DPI)

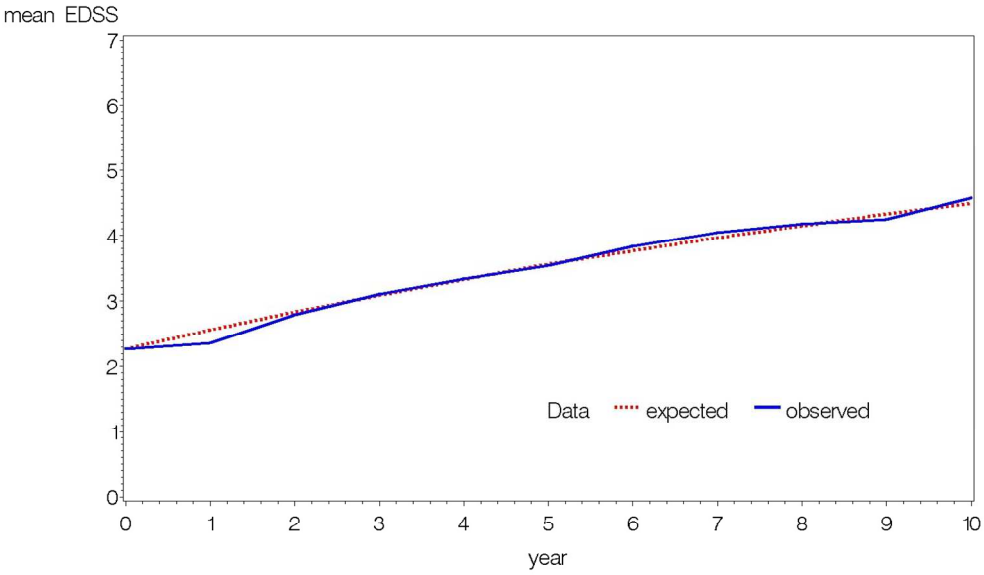
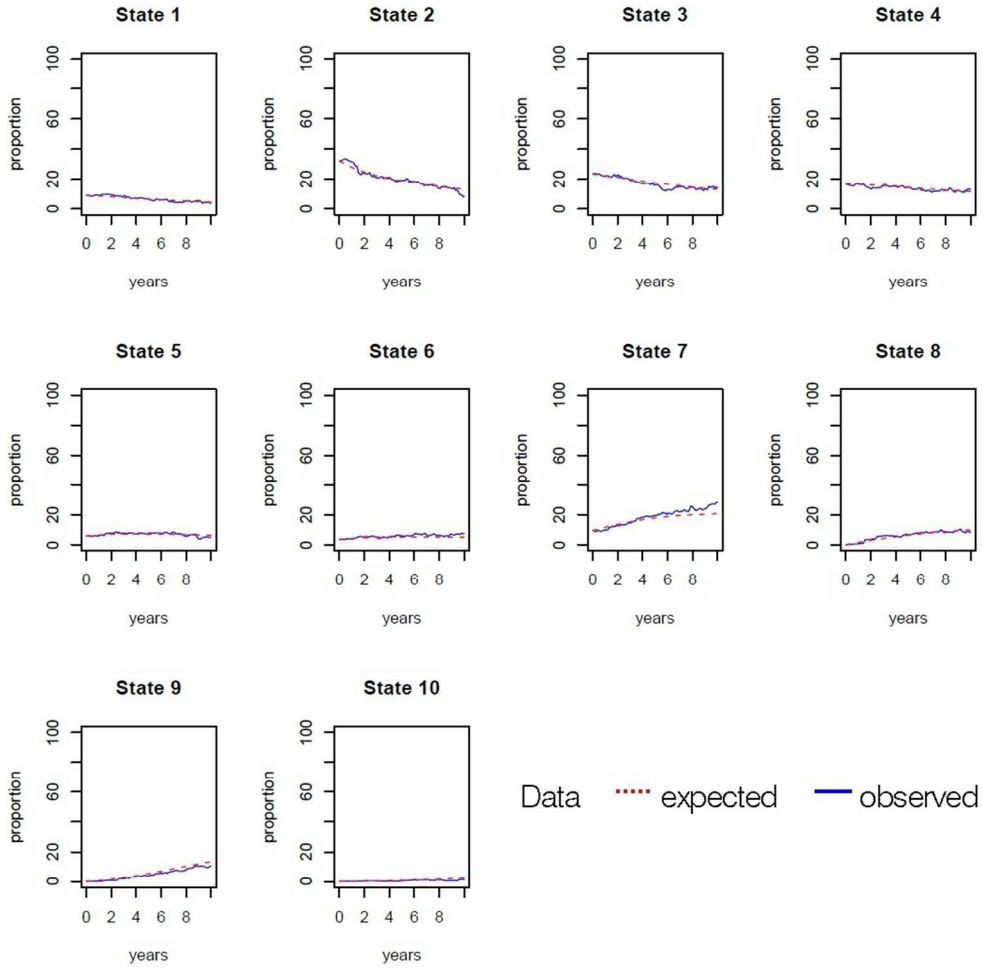


Figure 3: Transition probabilities obtained from the BCC dataset using the continuous Markov model with one covariate "age at onset" (binary version) were then applied to the baseline EDSS of the same cohort, projected over 10 years to produce a predicted outcome (red) compared to the observed course of the cohort (blue):
a) Mean EDSS shown in the predicted and actual cohorts
b)The proportion of patients predicted to be in each of the 10 EDSS states over time (state 1; EDSS 0, state 2; EDSS 1 and 1.5, state 3; EDSS 2.0 and 2.5, state 4; EDSS 3.0 and 3.5, state 5; EDSS 4.0 and 4.5, state 6; EDSS 5.0 and 5.5, state 7; EDSS 6.0 and 6.5, state 8; EDSS 7.0 and 7.5, state 9; EDSS 8.0 and 8.5, state 10; EDSS 9.0 and 9.5).

402x264mm (96 x 96 DPI)



283x280mm (96 x 96 DPI)

Appendix 1 EDSS data extraction from the BCMS database

For the discrete Markov model and for tabular display of annual data the EDSS data were extracted and processed as follows. If the baseline EDSS fell when it was not the 'usual' yearly visit period, then subsequent EDSS scores could be lost. To maximize the number of EDSS transitions per patient, but keeping the yearly (+/- three calendar months) interval, EDSS scores were also 'individualized' as follows: the baseline EDSS became the baseline year, within that year, we searched to find the optimal new, individualized baseline for which the patient would, over the coming years have the most number of yearly EDSS scores. For some patients, this new individualized baseline date would not coincide with a clinic visit and would therefore not have an EDSS score present. When a baseline EDSS was required (e.g. for the discrete Markov model), data was also 'shifted' such that each patient would have a baseline EDSS. This lag between baseline ('eligibility') and EDSS was considered consistent with clinical practice in that it is not unusual to have a lag time between a patient becoming eligible for treatment and actual treatment initiation.

Appendix 2 Algorithms to forecast EDSS distributions at any given time

The 'msm' algorithm allows the EDSS distribution to be forecasted at any time t.19 To define what is the actual EDSS at a given time t (i.e. not necessarily when an observation was recorded) 'msm' offered two variants: (i) the last observation carried forward (LOCF) for each individual patient, and (ii) a 'midpoint interpolation' algorithm in which the EDSS state for a given patient at time t was taken as the score closest in time to the actually observed EDSS. Suppose an individual was observed in EDSS states Sr-1 and Sr at two consecutive times tr-1 and tr, and we wanted to estimate ''observed' proportions at a time t between tr-1 and tr. LOCF then meant that individuals were assumed to be in state Sr-1 at time t, the same state as they were at tr-1. Midpoint interpolation meant if $t \leq (tr\text{-}1 + tr) / 2$, the midpoint of tr-1 and tr, the state at t was assumed to be Sr-1,, otherwise Sr. Option (i) would be more appropriate if EDSS values were always measured immediately after each transition. Option (ii) would be more appropriate if EDSS values were measured at mixed time intervals ('fixed or random'). Option (ii) was considered to mimic the clinical setting more closely and is also appears more applicable in progressive diseases.19 Therefore all results presented in this paper were based on this 'midpoint interpolation' approach.

For the continuous Markov model we limited the range of transitions which can be regarded as 'instantaneous' to +/- 3, which meant that at any time t an instantaneous progression (or improvement) into another EDSS state was only possible when not exceeding three consecutive states (an instantaneous transition from 1 to 2, 3, or 4 was possible, for example, whereas 1 to 5 was not). This restriction was recommended to avoid computationally inefficient modelling of hazard rates which were virtually 0.

Appendix 3 Validation of the models

The following validation techniques were applied when evaluating the different Markov models.

The most straightforward method for the discrete Markov model consisted of applying the transition matrix (and the 2nd, 3rd etc. power) to the vector of the baseline EDSS distribution in the BCMS reference database, calculating the forecasted EDSS distribution for t=1, 2, 3… years and comparing against the actual EDSS. A similar validation was performed in the continuous Markov model, as described by Jackson.1.

As a second validation we divided the BCMS dataset randomly into two subsets of equal size, using one half to derive the model separately for the two subgroups, and then assessing the goodness of fit in the other half of the dataset. Although Jackson19 emphasises that "Assessing the goodness of fit of this class of models […] is worth further research" we decided to use a classical mean square prediction error (weighted root mean square over years of the prediction error in the average quantity shown, weighted by the number of patients contributing data in the given year) to compare competing models. Moreover, the computed likelihood itself as a result of the maximum likelihood algorithm was used to rank the different one and two covariate models.