

Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator

Mike Boggild, consultant neurologist,¹ Jackie Palace, consultant neurologist,² Pelham Barton, senior lecturer in mathematical modelling,³ Yoav Ben-Shlomo, professor of clinical epidemiology,⁴ Thomas Bregenzer, director, biostatistics,⁵ Charles Dobson, senior projects officer,⁶ Richard Gray, director⁷

¹The Walton Centre, Liverpool L9 7LJ

²John Radcliffe Hospital, Oxford OX3 9DU

³Health Economics Unit, University of Birmingham, Birmingham B15 2TT

⁴Department of Social Medicine, Canynge Hall, Bristol BS8 2PS

⁵Biostatistics Unit, Parexel International, 14050 Berlin, Germany

⁶Department of Health, Quarry House, Leeds LS2 7UE

⁷University of Birmingham Clinical Trials Unit, Birmingham B15 2TT

Correspondence to: M Boggild mike.boggild@thewaltoncentre.nhs.uk

Cite this as: *BMJ* 2009;339:b4677 doi:10.1136/bmj.b4677

ABSTRACT

Objective To generate evidence on the longer term cost effectiveness of disease modifying treatments in patients with relapsing-remitting multiple sclerosis.

Design Prospective cohort study with historical comparator.

Setting Specialist multiple sclerosis clinics in 70 centres in the United Kingdom.

Participants Patients with relapsing-remitting multiple sclerosis who started treatment from May 2002 to April 2005 under the UK risk sharing scheme.

Interventions Treatment with interferon beta or glatiramer acetate in accordance with guidelines of the UK Association of British Neurologists.

Main outcome measures Observed utility weighted progression in disability at two years' follow-up assessed on the expanded disability status scale (EDSS) compared with that expected by applying the progression rates in a comparator dataset, modified for patients receiving treatment by multiplying by the hazard ratio derived separately for each disease modifying treatment from the randomised trials.

Results In the primary per protocol analysis, progression in disability was worse than that predicted and worse than that in the untreated comparator dataset ("deviation score" of 113%; excess in mean disability status scale 0.28). In sensitivity analyses, however, the deviation score varied from -72% (using raw baseline disability status scale scores, rather than applying a "no improvement" algorithm) to 156% (imputing missing data for year two from progression rates for year one).

Conclusions It is too early to reach any conclusion about the cost effectiveness of disease modifying treatments from this first interim analysis. Important methodological issues, including the need for additional comparator datasets, the potential bias from missing data, and the impact of the "no improvement" rule, will need to be addressed and long term follow-up of all patients is essential to secure meaningful results. Future analyses of the cohort are likely to be more informative, not least because they will be less sensitive to short term fluctuations in disability.

INTRODUCTION

In January 2002, the National Institute for Health and Clinical Excellence (NICE) published appraisal guidance on the use of disease modifying treatments for multiple sclerosis.¹ The drugs assessed were the beta interferons (Avonex, Betaferon, Rebif 22 µg, Rebif 44 µg) and glatiramer acetate (Copaxone). Randomised placebo controlled trials had shown the short term clinical effectiveness of each drug. To determine the cost utility of the long term effects of these treatments ScHARR (Sheffield School of Health and Related Research) created an economic model² using data on quality of life collected by the MS Trust from patients in the United Kingdom,³ cost data from Kobelt et al,⁴ a natural history dataset from London, Ontario,⁵ and estimates of delay in disease progression derived from randomised controlled trials. This model suggested that the disease modifying treatments were not cost effective over a 10 or 15 year horizon but became more cost effective over 20 years. It was uncertain, however, whether the results of randomised controlled trials lasting no more than three years could be reliably extrapolated over a longer period and whether patients who stopped treatment retained any benefits beyond that point, with disease progressing in line with that expected for untreated patients, or whether there was a rebound effect after cessation of treatment.

Given the above uncertainties, NICE was unable to recommend the treatments for use in the NHS at current prices but, instead, invited the Department of Health and the National Assembly for Wales to consider how they could be made available to patients in a cost effective manner.¹ In February 2002 the UK health departments set out the agreed basis of a "risk sharing scheme," which would allow the prescribing of Avonex, Betaferon, Copaxone, and Rebif 22 and 44 according to the Association of British Neurologists' 2001 guidelines,⁶ conditional on the development of a 10 year monitoring study that would collect data on the progression of disease in treated patients and thus help to assess the two critical uncertainties emphasised by NICE. If any individual product failed to show benefits consistent with projections made at the outset of the scheme, with the ScHARR model, the subsequent

price to the NHS would be reduced to restore cost effectiveness to a benchmark of £36 000 per quality adjusted life year (QALY) evaluated over a 20 year horizon. The choice of treatment was at the discretion of the neurologist in consultation with the patient. We report the results and implications of the first planned interim analysis, carried out when all patients had been in the scheme for at least two years.

METHODS

The broad principles for the establishment and conduct of the risk sharing scheme were set out in an NHS circular.⁷ The first patients who consented to take part in the study started treatment within two months of the publication of the circular. A steering committee was set up with representation from the parties to the scheme—that is, the health departments in England, Wales, Scotland, and Northern Ireland, each manufacturer, the MS Trust, the MS Society, the Association of British Neurologists, the Royal College of Nursing, and the UK MS Nurse Specialist Association.

The MS Trust acts as the secretariat to the scheme and custodian of the data. Two clinical neurology leads (JP, MB) provide clinical advice on the conduct of the study and liaise with other clinicians. For the three year recruitment phase SchARR coordinated the data collection and developed proposals for the statistical analysis in line with the principles set out in the circular. In 2006 a clinical research organisation (Parexel) took over data collection and statistical analysis, overseen by an independent scientific advisory group. The MS Trust, the MS Society, Parexel, the Department of Health, the four pharmaceutical companies, and the clinical leads attend scientific advisory group meetings (as observers). The remit of the group is to advise on scientific aspects of the study design, data analysis, and data interpretation and to assess additional study proposals; in particular, the scientific advisory group was responsible for advising on a detailed statistical analysis plan, amplifying the description of the planned primary analysis in the original health services circular.

Recruitment of patients

The circular suggested that 5000-7000 patients should be recruited to a monitoring cohort to allow about 1000 patients taking each treatment. Exploratory calculations suggested that this would be large enough to reduce sampling errors to an acceptable level. The 5583 patients assessed by clinicians as meeting the criteria of the Association of British Neurologists (broadly, patients with relapses as the predominant aspect of disease) were recruited into the monitoring study from May 2002 to April 2005 from 70 neurology centres across the UK. This risk sharing scheme cohort represents about 80% of patients with multiple sclerosis starting treatment in the UK over this period. Patients who entered the scheme with a diagnosis of secondary progressive multiple sclerosis were analysed separately. The main analyses in this paper relate

to the 4749 patients with relapsing-remitting multiple sclerosis at entry.

Data collection

As this was a pragmatic study undertaken within routine practice, it was designed to collect minimum basic information through clinical assessments with expanded disability status scale (EDSS) scores⁸ performed annually within a three month window of the entry date into the study. Patients were to be followed even if treatment was stopped or altered to minimise “dropout bias” and to test the SchARR model assumptions that benefit up to that point is maintained over the longer term.

Outcome measures

With no randomised control group, we had to compare the disease progression for the risk sharing cohort with a cohort of patients who were recruited and followed up before the routine use of disease modifying treatments. For the purposes both of the original SchARR model and of the risk sharing scheme, the London, Ontario, dataset was considered to be the largest, most complete population based source of such data.⁵ This dataset consists of 1043 Canadian patients, recruited in 1972-84, whose disability was assessed about annually for a median of 25 years from onset of disease with the disability status scale (DSS, the predecessor of the EDSS). As in the risk sharing cohort, in the Ontario dataset diagnoses were made by experienced neurologists according to standard clinical criteria.⁹ Although patients in this cohort were followed largely prospectively, the dataset derived from this seems to have been collated retrospectively with the disability status scores smoothed to eliminate short term fluctuations. As a result there are no “regressions” in EDSS in the Ontario dataset; disability scores for individual patients can only worsen over time and the SchARR model had to reflect this.

For each treatment separately we used the SchARR model to predict the expected movement of patients between the EDSS states both “on” and “off” treatment. For patients “off” treatment, the SchARR model uses a matrix of transition probabilities derived from the actual progressions seen in the Ontario dataset (using a subset of 314 patients who were judged to have fulfilled the Association of British Neurologists’ criteria—experiencing two or more relapses in the previous two years—at baseline). These transition matrices are modified for patients “on” treatment by multiplying by the hazard ratio (relative rate of disease progression) derived separately for each treatment from the pivotal 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 appropriate subset of patients in the risk sharing scheme cohort. As in SchARR’s original work for the NICE appraisal, we assumed that patients who stop taking treatment will experience the same rate of disease progression from that point onwards as patients at the same EDSS point who have never been treated.

Our primary outcome measure was a deviation score of the average observed loss of utility (average utility weighted disease progression; see appendix 1 on bmj.com) for patients in the risk sharing scheme compared with the expected loss calculated by the SchHARR model for patients “on” treatment. Utilities were derived from a two stage survey of 1554 respondents from the MS Trust database (78% response rate to second questionnaire, 18% of total database) who completed the EQ-5D, which was then converted to a single utility, the EQ-5D index score.³

We calculated the primary outcome from several other measures. The “expected benefit” of treatment (with a specific treatment) is the “hypothetical” difference between the expected outcome without and with treatment, as calculated in each case from the SchHARR 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 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 implies that it was worse than predicted, and a value of 0 indicates that it was exactly as predicted. This deviation measure is calculated every two years and used as the basis of possible price adjustments under the agreed rules of the scheme; thus if the shortfall between actual and expected benefit exceeds an agreed “tolerance margin” (20% at year two, 10% at subsequent review points) the Department of Health and relevant manufacturer will renegotiate the current price for the treatment so that it achieves cost effectiveness at £36 000 with the new estimated treatment effect. We have also reported changes in the average (unweighted) EDSS score, comparing the expected and actual change from baseline, as this will be more familiar and easier to conceptualise.

Statistical methods

The primary analysis plan made maximum use of the available data, including data on patients with only one year of valid follow-up. For this subset, comparison is with the expected progression over one year as calculated by the SchHARR model.

One major assumption of the SchHARR model (following the conventions of the Ontario dataset) is that EDSS scores are constrained to remain stable or worsen and improvements are not possible. We therefore had to apply an algorithm to the risk sharing scheme cohort to model as closely as possible the way in which we understand the Ontario dataset to have been compiled (we did not have access to the raw Ontario data). This algorithm (see appendix 2 on bmj.com) includes, for some patients, a (downwards) adjustment to the baseline EDSS score where the “raw” data would otherwise have implied an improvement from baseline to year one or in subsequent years. In addition, the algorithm means that an apparent disease progression to year two, for example, is disregarded if it is not

confirmed by the data for year three. For many patients in the analysis dataset for the risk sharing scheme, however, three year assessments are not yet available and so, in the primary analyses, we accepted unconfirmed disease progressions at this stage, even though further follow-up might modify these.

The primary (per protocol) analysis censored patients who died from causes other than multiple sclerosis, emigrated, were lost to follow-up, or switched treatments. Patients who died from a cause related to multiple sclerosis had their subsequent missing EDSS score altered to a value of 10. The SchHARR model assumed that all those who progressed to secondary progressive multiple sclerosis would stop treatment and could be treated in the same way as other patients who stopped treatment (that is, assuming the same rate of disease progression as untreated patients). Patients who moved from relapsing-remitting multiple sclerosis to secondary progressive multiple sclerosis but who still took treatment (regardless of whether the treatment was actually licensed for use in secondary progressive multiple sclerosis or not) were censored after the first assessment after conversion and were then treated as “lost to follow-up” because in the original scheme we did not expect that there would be many such patients and had not defined a clear basis for estimating their “expected” disease progression. We have, however, carried out a sensitivity analysis to include these patients.

The scientific advisory group recognised that some of the exclusions or censoring would probably bias the progression data. Also, at this early stage, the results could be unduly influenced by a small number of patients with extreme changes in EDSS score (in either direction). The scientific advisory group, therefore, suggested several supplementary sensitivity analyses (table 1). Some of these were specified in advance, others in the light of unexpected features in the year two data.

RESULTS

Participants

Out of 5583 patients registered into the monitoring scheme, 4749 (85%) had relapsing-remitting multiple sclerosis. Of these, 208 (4.4%) were ineligible for analysis and we excluded 248 (5.2%) because they were never treated or did not start treatment within three months of baseline assessment (fig 1). Of the 4293 who were eligible and treated, 358 (8.3%) had no subsequent follow-up data and a further 249 (5.8%) had data that were not used in the primary analysis, mostly because the patient had switched treatment or had converted to secondary progressive multiple sclerosis in the previous year and remained taking treatment (fig 1). Thus 3686 (85.9%) of eligible and treated patients had at least some valid follow-up data. These 3686 patients are referred to as the “per protocol analysis cohort.” Of these, 2901/4293 (67.6%) had valid EDSS assessments at year two and 785 patients had valid EDSS assessment at year one but not at year two. Of the 2850 patients who were still relapsing-

Table 1 | List of sensitivity analyses undertaken with rationale

Problem to be addressed	Sensitivity analysis
Patients switching treatment might have particularly poor outcomes so excluding them might bias results	Calculate on "intent to treat" basis, retaining all such patients in analysis
Patients converting from RRMS to SPMS but remaining on DMTs might have particularly aggressive disease, so excluding them might bias results	Retain patients in analysis after developing SPMS and still taking DMTs; calculate "expected" benefit on assumption that DMTs modify natural rate of disease progression by same factor as shown for RRMS in randomised controlled trials*
Results might be distorted by small number of outliers	Exclude patients with 1% most extreme movement (up or down) in EDSS score
Bias because of lack of year three data for some patients needed to confirm apparent progression at year two	Limit analysis to subset of patients with data for year three
Bias because of lack of year three data for some patients needed to confirm apparent progression at year two	Use all available data to estimate proportion of apparent progressions subsequently confirmed and apply this to patients with apparent progression at year two but no year three data*
Possible bias because of missing year two data (see below)	Perform "best case/worst case" analysis to impute missing year two values—for "best case" assume no further disease progression after year one, for "worst case" extrapolate from progression between baseline and year one
Possible bias because of retrospective adjustment of baseline data for some patients	Leave baseline EDSS values unadjusted while continuing to apply "smoothing" algorithm to subsequent data points

RRMS=relapsing-remitting multiple sclerosis; SPMS=secondary progressive multiple sclerosis; DMTs=disease modifying treatments; EDSS=expanded disability status scale.
*Pre-specified analysis

remitting and had valid EDSS assessments at year two, 2609 (91.5%) were still receiving their original disease modifying treatment. An additional 228 patients with relapsing-remitting multiple sclerosis had switched treatment before their first available assessment. Patients converting to secondary progressive multiple

sclerosis who stopped treatment were included in the per protocol cohort, but, by the end of the two year period, only 48 (12.0%) of 401 patients who progressed to secondary progressive multiple sclerosis had discontinued treatment by their next annual assessment.

Table 2 shows the baseline data characteristics of all the eligible patients with relapsing-remitting multiple sclerosis at baseline and the subset used in the primary analysis. The age and sex distributions, the baseline EDSS values, previous relapse rates, and times from symptom onset and diagnosis are almost identical for the two groups. A comparison with data for the patients with secondary progressive multiple sclerosis at baseline shows that, as expected, the latter have higher baseline EDSS values and a longer time from symptom onset and diagnosis (data not shown).

Over the two year period 1403 (38%) patients in the per protocol analysis set showed an improvement in EDSS scores, and for 591 patients (16%) this was confirmed either up to year two or at the next annual assessment. These proportions are in contrast with the assumption in the SchARR model, based on the Ontario dataset, that improvements (especially sustained improvements) in annual scores are unlikely. Over the same period 1803 (49%) patients had deterioration in EDSS, and in 834 (23%) this was confirmed. At year two, 2629 patients (71.3% of patients in the per protocol analysis cohort) were still taking their initial treatment, but 272 (7.4%) had stopped all treatment, 214 (5.8%) had switched to a different disease modifying treatment, and 571 (15.5%) had become lost to follow-up, died, or had missing EDSS data at year two. For the per protocol analysis fig 2 shows a comparison of expected and actual EDSS at follow-up.

The mean annual rate of change in the EDSS score in the per protocol analysis cohort was 0.35 after application of the "no improvement" rules and 0.16 with the "raw" data. As a result of the no improvement rules, of the 3686 patients in the per protocol analysis cohort, we modified EDSS scores of 989 (26.8%) patients at

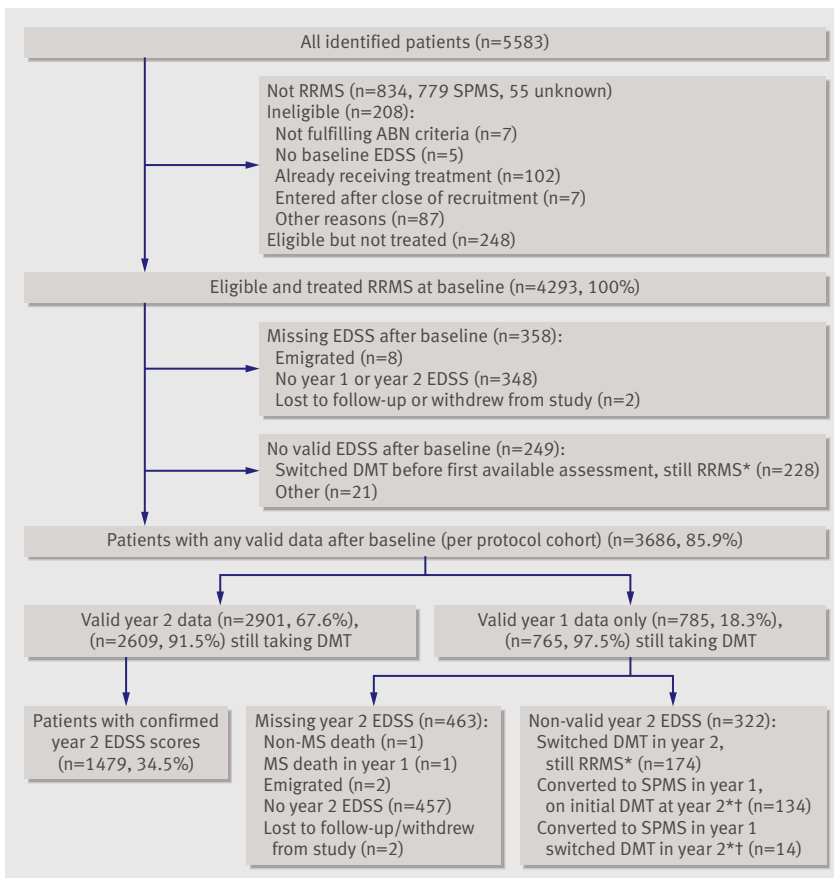


Fig 1 | Participants with relapsing-remitting multiple sclerosis (RRMS) taking disease modifying treatments (DMT) in multiple sclerosis risk sharing scheme. *Data used for supplementary analyses. †Patients assessed as secondary progressive multiple sclerosis (SPMS) at year 1 follow-up; their EDSS scores for year 1 (but not for later years) retained in primary analysis

Table 2 | Baseline characteristics of all eligible and treated patients with relapsing-remitting multiple sclerosis (RRMS) at entry and those included in per protocol analysis cohort

Characteristic	All eligible and treated patients with RRMS at entry (n=4293*)	Per protocol analysis cohort (n=3686†)
No (%) of men	1060 (24.7)	906 (24.6)
No (%) of women	3233 (75.3)	2780 (75.4)
Median (range) age (years)	38 (18-73)	38 (18-73)
Mean (SD, range) EDSS	3.07 (1.52, 0.0-6.5)	3.05 (1.52, 0-6.5)
Median (range) time since symptom onset (years)	5.7 (0-48)	5.7 (0-48)
Median (range) time since first diagnosis (years)	2.6 (0-41)	2.6 (0-41)
Median (range) No of confirmed relapses in last 2 years	3 (0-21)	3 (0-21)

*Missing data: 5 for age, 20 for symptom onset, 41 for time since diagnosis, 3 for previous relapses.
†Missing data: 5 for age, 16 for symptom onset, 36 for time since diagnosis, 2 for previous relapses.

baseline and 715 patients (19.4%) for at least one follow-up visit.

The baseline EDSS and mean change per year for the primary outcome cohort compared with those patients who have only one year of valid follow-up data show that the latter subgroup were more disabled at baseline and progressed more rapidly. Thus mean EDSS for the 2901 patients in the per protocol analysis with valid EDSS data at year two was 2.68 at baseline, 2.90 after one year, and 3.24 after two years. Patients who had only one year of valid data (n=785) had mean EDSS of 3.14 at baseline and 3.77 after one year (P<0.001 for comparison of baseline and of rate of change, Mann-Whitney-Wilcoxon test).

Primary analysis

The primary analysis shows a positive deviation score of 113% for the weighted utility score, the primary outcome measure (table 3). At face value this indicates that cost effectiveness of disease modifying treatments was worse than that expected from the ScHARR model as applied to the Ontario dataset and also worse than that predicted with no treatment from the same model. In absolute terms the EDSS score was 0.10 unit worse than the control data and 0.28 units worse than predicted on the assumption that the disease modifying treatments delay progression of the disease.

Sensitivity analyses

Retaining all patients in the analysis regardless of any switching between treatments, and retaining patients

who moved from relapsing-remitting to secondary progressive multiple sclerosis but remained taking treatment, made virtually no difference to the deviation measure. The difference between observed and expected EDSS progression was less marked when we restricted the analysis to the subgroup of patients with the year two EDSS score confirmed by valid year three data (0.14 v 0.28) or when we adjusted the results to take account of the expected proportion of apparent progressions at year two that were not subsequently confirmed at year three (0.19 v 0.28). Imputation of data for year two in patients with only year one values gave results marginally more favourable to the disease modifying treatments in the “best case” scenario (0.27) but distinctly worse in the “worst case” scenario (0.33) compared with the primary analysis (0.28). When we used the unadjusted EDSS scores at baseline (thereby allowing improvement from baseline to year one) while continuing to apply the “no improvement” algorithm to subsequent data points, however, there was a negative deviation measure (-84)—that is, patients progressed less rapidly than predicted from the ScHARR model (-0.11 EDSS points compared with 0.28 in the per protocol analysis) (fig 3).

DISCUSSION

In this observational cohort study of a risk sharing scheme for disease modifying treatments in patients with multiple sclerosis we found no evidence that these treatments are cost effective. The purpose of the scheme is to provide patients with access to treatments and at the same time collect evidence to help to assess their cost effectiveness in a standard NHS setting for future pricing negotiations.

The outcomes so far obtained in the pre-specified primary analysis suggest a lack of delay in disease progression for all disease modifying treatments combined. Some of the sensitivity analyses performed were more favourable to the drugs, with one even implying better outcomes than expected on the ScHARR model, although one (the “worst case” version of the analysis addressing possible bias because of missing year two data) gave less favourable results than the primary analysis, suggesting a bias in favour of treatment resulting from missing data.

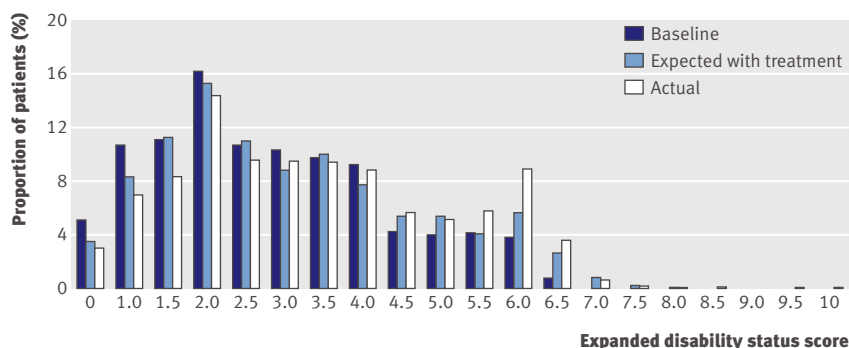
**Fig 2** | Expected and observed EDSS at follow-up for per protocol analysis

Table 3 | Primary outcome and sensitivity analyses with EDSS and deviation score

	Predicted progression from transition probabilities		Actual progression observed with treatment	Difference (actual less predicted)	Deviation score (%)
	Without treatment	With treatment			
Per protocol analysis (n=3686)					
EDSS	0.47	0.29	0.57 (0.544 to 0.603)	0.28	—
Utility score	0.0254	0.0158	0.0266	—	113
Including DMT switchers (n=3931)					
EDSS	0.48	0.30	0.59 (0.559 to 0.617)	0.28	—
Utility score	0.0258	0.0162	0.0271	—	113
Including SPMS converters on DMTs (n=3690)					
EDSS	0.48	0.30	0.58 (0.550 to 0.610)	0.27	—
Utility score	0.0261	0.0164	0.0275	—	114
Excluding extreme outliers (n=3650)					
EDSS	0.47	0.30	0.54 (0.514 to 0.570)	0.25	—
Utility score	0.0254	0.0158	0.0250	—	96
Subgroup analysis with year 2 confirmed scores (n=1479)					
EDSS	0.50	0.32	0.46 (0.422 to 0.507)	0.14	—
Utility score	0.0263	0.0162	0.0214	—	51
Predicted change in unconfirmed year 2 scores (n=3686)					
EDSS	0.47	0.29	0.48*	0.19	—
Utility score	0.0254	0.0158	0.0225	—	70
Imputing missing year 2 scores—best case scenario (n=3686)					
EDSS	0.53	0.33	0.60 (0.566 to 0.627)	0.27	—
Utility score	0.0290	0.0179	0.0285	—	95
Imputing missing year 2 scores—worst case scenario (n=3686)					
EDSS	0.53	0.33	0.66 (0.629 to 0.699)	0.33	—
Utility score	0.0290	0.0179	0.0351	—	156
Unadjusted baseline scores (n=3686)					
EDSS	0.49	0.31	0.20 (0.163 to 0.241)	-0.11	—
Utility score	0.0257	0.0157	0.0073	—	-84

DMTs=disease modifying treatments; EDSS=expanded disability status scale.
 *95% confidence intervals cannot be calculated by standard methods because of indirect method used to estimate mean EDSS change.

Factors to consider

Our use of the no improvement “rules” might have underestimated the effects of treatment. As already noted, the ScHARR model does not permit regressions (that is, improvement in disability) and the data from the risk sharing scheme had to be modified to allow comparison with the predicted disease progression. EDSS scores for patients with multiple sclerosis, however, do spontaneously improve for various reasons: recovery from relapse, natural fluctuations, and measurement error. In addition, it is clear that the disease modifying treatments do reduce relapses¹⁰ and might allow improved recovery from relapse early in treatment. In the risk sharing cohort, 32% of patients show some improvement in disability scores and for 14% this improvement is sustained for a second year.

To maintain comparability with the Ontario dataset, we did not use the raw EDSS scores as this would have biased the deviance score to overestimate the true benefits of treatment. It is less clear whether the primary analysis is correct in applying the no improvement algorithm to adjust the baseline score as well as the subsequent scores; using the unadjusted baseline values had a marked effect on the results. The impact of the “no improvement” assumption will diminish with longer follow-up as patients will show greater

change in their EDSS scores and the effects of any minor misclassifications early in follow-up will be of less significance, but it will remain a source of some uncertainty in estimating cost effectiveness.

The Ontario comparator dataset might not be appropriate for other reasons, such as the natural course of the disease changing over time. This would be a valid explanation only if patients progressed more rapidly today than 20 years ago. Data suggest that patients with multiple sclerosis now progress more slowly,¹¹ so comparison with historical data should overestimate the potential benefits of treatment. Compared with the primary out-

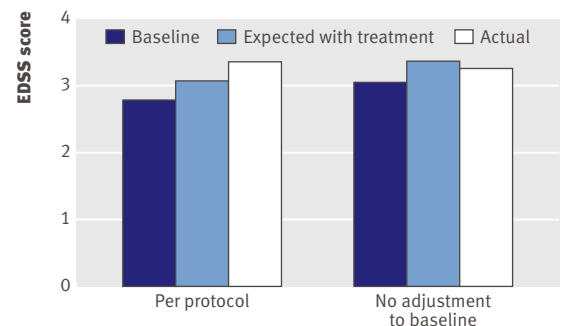


Fig 3 | Effect of baseline adjustment of EDSS

come cohort the Ontario dataset differs in other covariates, such as age at presentation and sex, that are associated with disease progression. We have not explored this possibility as we do not have access to the raw Ontario data, but the transition probabilities for the Ontario cohort were taken for a subgroup of patients fulfilling the Association of British Neurologists' criteria for inclusion in the multiple sclerosis risk sharing scheme to make it as comparable as possible.

These uncertainties make reliable interpretation of the short term results problematic, and we have not presented data for the individual treatments, which are likely to be further confounded by selection bias because disease severity might itself determine which treatment is used. The scientific advisory group considered that it was premature, at this stage, to reach any decision about re-pricing the drugs without further follow-up and analyses.

Limitations and future plans

Estimating cost effectiveness from an observational cohort study with a historical comparator dataset has inherent problems that would be avoided in a randomised controlled trial.¹² In 2002, however, a placebo controlled study lasting for 10 years or more was not deemed feasible as the safety and efficacy of these products had already been accepted by the regulatory authorities. In addition, patients would be unwilling to be randomised to placebo for such a long period, the dropout rate would have been high, and some clinicians who consider that the existing evidence on efficacy is already convincing would have regarded allocation to placebo as unethical. With hindsight the inclusion of further sensitivity analyses in pivotal publications could have shed additional light on the uncertainties about treatments in multiple sclerosis.^{10 12}

We appreciated from the outset that a major limitation was the validity and generalisability of the comparison dataset. Consistent results with various different comparator databases would add confidence to conclusions about the cost effectiveness of treatment. The scientific advisory group is actively pursuing alternative sources of data on disease progression in untreated patients. This would have two major benefits. Firstly, it would allow estimates of progression rates based on unadjusted data where EDSS scores can get better as well as worse and hence the analysis can be repeated on the raw rather than adjusted scores. Secondly, it would allow examination of whether the rates of disease progression shown in the Ontario cohort are similar, at least in the short term, to those in other cohorts from different geographical as well as temporal populations of patients.

Some of the patients who entered the risk sharing scheme (and therefore started disease modifying treatments) at the outset will have had disease for a longer duration than is current clinical practice, as access to treatment in the UK was fairly restricted before the introduction of the scheme. This would potentially bias the results to underestimate the effects of treatment if (as recent results suggest) such patients show

less benefit than those treated earlier in the natural course.¹³ Another potential bias, which would overestimate the benefits of treatment, is incomplete follow-up. Completeness of follow-up might well differ between the treated and reference cohorts. As expected, patients with incomplete follow-up data in the risk sharing scheme had more rapid disease progression: one year decline was greater in those with one year but not two year assessments than in those who were assessed at both time points. A best case-worst case analysis suggested that adjustment for this effect would probably give even less favourable results than the per protocol analysis. Conversely, when we applied a correction to allow for some of the year two progressions not being confirmed by the subsequent year three results we saw a smaller positive deviation score. Imputing data might be problematic and bias caused by loss to follow-up is a major concern in such a long term study, highlighting the importance of obtaining data from all patients who took part in the scheme even if they are no longer receiving treatment. Clinic assessments are not always possible—for example, patients who have reached high disability scores (EDSS ≥ 7.0) and stopped treatment because of progression can have difficulty attending the clinic. For this reason, an option is being introduced of having a telephone assessment of EDSS undertaken by a multiple sclerosis nurse. The telephone EDSS by doctors has been validated in a pan-European study¹⁴ and a parallel validation study by multiple sclerosis nurses is currently under way.

One of the uncertainties in the original SchARR model was the estimate of costs and utilities. If the cost of problems related to multiple sclerosis is not fully captured then the cost effectiveness of treatment will be underestimated. An additional study to capture cost and utility data has therefore been undertaken and will be reported on shortly to help inform the year four analysis.

The original analysis plan wrongly assumed that patients who developed secondary progressive multiple sclerosis would stop treatment and would thus be modelled using the untreated transition probabilities. Most such patients in the scheme, however, continued treatments whether or not they were licensed for secondary progressive multiple sclerosis. There are several possible reasons. Firstly, all the treatments probably reduce relapses, whatever the stage of the disease. Secondly, withdrawal of treatment might be deferred until the clinician and patient are sure that secondary progression has occurred, which might not be before one or even two annual cycles of the scheme. Finally, as there are limited further treatment options patients can be resistant to withdrawal of treatment. Our secondary analysis showed, perhaps slightly against expectation, that including patients who developed secondary progressive multiple sclerosis while taking treatment did not have a significant impact on the results.

Since the onset of the risk sharing scheme, there have been changes in the management of patients with multiple sclerosis. The updated guidelines from the

Association of British Neurologists¹⁵ have widened the eligibility criteria for disease modifying treatments, though funding has not been agreed with the Department of Health. This itself will not impact on the results of the scheme, although it might make it more difficult to generalise the findings to current practice. With the licensing of new classes of drug such as natalizumab, patients whose relapses do not respond to the scheme drugs (either because they have an aggressive form of multiple sclerosis or because they have neutralising antibodies to interferon beta) might be switched to newer non-scheme drugs. To avoid bias, it is important that such patients are included in the analysis, even if only in sensitivity analyses.

Further efforts are being made to trace patients with missing data at year two, a disproportionate number of whom might have stopped taking treatment because of more aggressive disease. This includes the treating clinicians assessing disability by telephone and individually reviewing case notes for those with missing data. In addition, it would have been desirable to have identified patients who were eligible for the scheme but chose not to participate or who were recruited but then decided not to have treatment. It would then have been possible to use newer statistical methods that allow less biased estimates of the potential causal effects from observational data, as has been done with antiretroviral therapy for HIV progression.¹⁶

Risk sharing schemes

The risk sharing scheme was an innovative solution to a familiar dilemma in health technology assessment—the problem of assessing the longer term benefits of treatments for progressive diseases when, because of cost considerations and the difficulties in maintaining large populations in prolonged placebo controlled studies, pivotal clinical trials maintain randomisation for only a relatively short period compared with the natural course of the disease.

The establishment of the risk sharing scheme for disease modifying treatments in multiple sclerosis has allowed thousands of patients to have access to certain multiple sclerosis drugs, which they might not have had if the NICE evaluation had been implemented, and has been the catalyst for major changes in the management of multiple sclerosis, including substantial increases in the number of centres with specialist multi-disciplinary teams and in the number of specialist nurses. From the perspective of patients, these are all positive developments. This has been achieved by use of NHS resources, which would otherwise have been available for the treatment of other patients. Whether the gain to patients with multiple sclerosis exceeds the loss to those unidentifiable other individuals remains unresolved.

Other benefits include the establishment of a large cohort of patients that could be used for supplementary studies on risk factors for disease progression, the development of expertise at over 70 special multiple sclerosis treatment centres, and increased funding for specialist nurses so that now over 200 are employed

compared with 75 at the start of the scheme. While acknowledging these indirect effects of the establishment of the risk sharing scheme, it should be noted that the direct costs of running the scheme over a 10 year period will represent only around 1% of the costs to the NHS of the disease modifying treatments over this period.

A unique element of this scheme was that the Department of Health and drug companies negotiated a reduction in price of some treatments at the outset to achieve what was thought to be the anticipated level of cost effectiveness. This practice has been pioneered in Australia, where the Pharmaceutical Benefits Advisory Committee often negotiates the price of a drug downwards. The Pharmaceutical Price Regulation Scheme (PPRS), which governs drug pricing in the UK, has not used this approach in the past, although the most recent revision to the scheme encourages companies to propose a cost effective price on first launching their products in the UK market, based on the evidence then available, and allows them to propose a subsequent price increase in the light of any further evidence on cost effectiveness.¹⁷

Other forms of “risk sharing” have emerged since this scheme was established, but most relate to establishing an “outcomes guarantee”¹⁸ for treatments in which the benefit to individual patients can be clinically assessed. An example of this is bortezomib (Velcade) for the treatment of relapsed multiple myeloma, where a “response rebate” scheme was established by the Department of Health so that patients who showed no or minimal response are taken off treatment and drug costs are refunded by the manufacturer.¹⁹

The multiple sclerosis risk sharing scheme could be better termed as “coverage with evidence development,” which has been used in the past few years in countries such as the United States, Australia, Canada, and Europe. This provides interim approval for reimbursement of a treatment, conditional on the collection of further evidence and review after a specified period.²⁰ This process allows patients’ access to promising new treatments but manages that access in a coordinated way, generating additional evidence that is targeted to reduce uncertainties in a much more structured way than a traditional post-marketing study. While this sounds an attractive option, the comparison of contemporary observational data with historical cohorts is notoriously problematic. For example, the rate of progression in the placebo arm of a screening evaluation of potential neuroprotective treatments for Parkinson’s disease was significantly lower than that in the historical reference cohort and thus placebo treatment met the criteria for being taken forward as a promising neuroprotectant.²¹

Summary

The UK multiple sclerosis risk sharing scheme was an innovative approach to collect new cost effectiveness data as part of an observational clinical cohort study. The two year results highlight many of the methodological difficulties of such an approach, including the

WHAT IS ALREADY KNOWN ON THIS TOPIC

Randomised controlled trials have shown that disease modifying treatments can slow the progression of relapsing-remitting multiple sclerosis over a two to three year period

It is not known whether these benefits persist over the longer term or whether patients who stop treatment retain the benefit they have received up to that point

For these reasons, it is not clear whether disease modifying treatments represent a cost effective use of NHS resources

WHAT THIS STUDY ADDS

This two year interim analysis of the UK risk sharing scheme does not provide reliable evidence on cost effectiveness of disease modifying treatments

Results are highly sensitive to how the baseline score and missing follow-up data are handled

Longer more complete follow-up with additional reference datasets will be more informative and less sensitive to short term fluctuations in disability

inherent difficulty with historical comparators, and the major uncertainties around the interpretation of the current results. The primary analysis did not meet the predefined level for cost effectiveness but, at this stage, we cannot reliably determine whether the current pricing of these drugs represents value for money for the NHS. Longer term follow-up will reduce some of the uncertainties arising from short term fluctuations in disability scores and will provide new empirical evidence to confirm or refute some of the assumptions made by the NICE committee when considering the cost effectiveness of disease modifying treatments. The longer term success of the scheme, however, is dependent on the hard work and goodwill of many NHS staff in continuing to collect data on all patients who were entered into the scheme.

We acknowledge the contribution of Cindy Cooper and Mark Pickin in relation to patient recruitment to the scheme

Contributors: CD was involved in the original negotiation of the risk sharing scheme and continues to advise the parties to the scheme on analytical aspects. MB and JP are the clinical leads for the project. MB, JP, PB, YB-S, TB, and RG helped design the primary analysis plan, interpreted the results, and suggested further sensitivity analyses. TB ran the analyses. All the authors have contributed to the final version of the paper. MB is guarantor.

Funding: The central costs of data collection and analysis were shared in five equal amounts between the Department of Health and the four pharmaceutical companies (Bayer Schering Pharma, Biogen, Merck Serono, Teva). Treatment costs (drug and staff) were largely met by local commissioners, with financial support for nursing and other staff from the companies and the MS Society. Analysis and interpretation of results and drafting of the manuscript were undertaken by a group independent of the study sponsors.

Competing interests: MB and JP have received support for attending international congresses from each of the four pharmaceutical companies funding the study. CD is an employee of the Department of Health and was involved in the negotiation of the original risk sharing scheme.

Ethical approval: This study was approved by the South East multicentre research ethics committee (02/10/78).

Data sharing: Data from the scheme may be available for collaborative studies, applications should be made to the scientific advisory group via the clinical leads (JP, MB).

- 1 National Institute for Health and Clinical Excellence. Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. NICE, 2002 (Technology Appraisal Guidance No 32).
- 2 Chilcott J, McCabe C, Tappenden P, Cooper NJ, Abrams K, Claxton K. Modelling the cost-effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. *BMJ* 2003;326:522-6.
- 3 Hemmett L, Holmes J, Barnes M, Russell N. What drives quality of life in multiple sclerosis? *QJM* 2004;97:671-6.
- 4 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>).
- 5 Ebers G. Natural history of multiple sclerosis. In: Compston A, Ebers G, Lassmann H, McDonald IR, Matthews B, Wekerle H. *McAlpine's multiple sclerosis*. 3rd ed. Churchill Livingstone, 1998.
- 6 Association of British Neurologists. Guidelines for treatment with interferon beta and glatiramer acetate in multiple sclerosis. ABN, 2001. www.theabn.org/documents/msdoc.pdf.
- 7 Department of Health. Cost effective provision of disease modifying therapies for people with multiple sclerosis. Health Service Circular, 2002/004 (latest update October 2006). www.dh.gov.uk/en/PublicationsAndStatistics/LettersAndCirculars/HealthServiceCirculars/DH_4004332.
- 8 Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale. *Neurology* 1983;33:1444-52.
- 9 Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis F, Ebers GE, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;13:227-31.
- 10 Rice G, Inorvaia B, Munari L, Ebers G, Polman C, D'Amico R, et al. Interferon in relapsing-remitting multiple sclerosis. *Cochrane Database Syst Rev* 2001;(4):CD002002.
- 11 Tremlett H, Paty D, Devonshire V. Disability progression in multiple sclerosis is slower than previously reported. *Neurology* 2006;24:172-7.
- 12 Sudlow CLM, Counsell CE. Problems with UK government's risk sharing scheme for assessing drugs for multiple sclerosis. *BMJ* 2003;326:388-92.
- 13 Kappos L, Freedman MS, Polman CH, Edan G, Hartung HP, Miller DH, et al. Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. *Lancet* 2007;370:389-97.
- 14 Lechner-Scott J, Kappos L, Hofman M, Polman C, Ronner H, Montalban X, et al. Can the expanded disability status scale be assessed by telephone? *Multiple Sclerosis* 2003;9:154-9.
- 15 Association of British Neurologists. Guidelines for treatment of multiple sclerosis with interferon beta and glatiramer acetate. ABN, 2007.
- 16 Sterne J, Hernán M, Ledergerber B, Tilling K, Weber R, Sendi P, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet* 2005;366:378-84.
- 17 Department of Health/ABPI. The 2009 pharmaceutical price regulation scheme. www.dh.gov.uk/en/Publicationsandstatistics/Publications/DH_091825.
- 18 Chapman S, Reeve E, Rajaratnam G, Neary R. Setting up an outcomes guarantee for pharmaceuticals: new approach to risk sharing in primary care. *BMJ* 2003;326:707-9.
- 19 National Institute for Health and Clinical Excellence. Bortezomib monotherapy for relapsed multiple myeloma. NICE, 2007 (Technology Appraisal Guidance No 129).
- 20 Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *Int J Tech Ass Health Care* 2007;23:425-35.
- 21 NINDS NET-PD Investigators. A randomised, double-blind, futility clinical trial of creatine and minocycline in early Parkinson disease. *Neurology* 2006;66:664-71.

Accepted: 5 August 2009