Evidence Review Group's Report
Teriflunomide for treating relapsing forms of multiple sclerosis

Produced by CRD and CHE Technology Assessment Group, University of York, Heslington, York, YO10 5DD

Authors Debra Fayter, Research Fellow, CRD

Eldon Spackman, Research Fellow, CHE

David Epstein, Research Fellow, CHE

Stephen Palmer, Professor of Health Economics, CHE

Rocio Rodriguez-Lopez, Information Specialist, CRD

Nerys Woolacott, Senior Research Fellow, CRD

Correspondence to Dr Nerys Woolacott, CRD, University of York, York YO10 5DD

Date completed 29/07/2013

Source of funding

This report was commissioned by the NIHR HTA Programme as project number HTA/12/06.

Declared competing interests of the authors

None

Acknowledgements

We thank Paul Talbot, Department of Neurology, Salford Royal NHS Foundation Trust Hope Hospital, Stott Lane, Salford M6 8HD who acted as clinical advisor to the ERG.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Fayter D, Spackman E, Epstein D, Palmer S, Rodriguez-Lopez R, Woolacott N: Teriflunomide for treating relapsing forms of multiple sclerosis. A Single Technology Appraisal. CRD and CHE Technology Assessment Group, 2013.

Contributions of authors

Debra Fayter and Nerys Woolacott wrote the clinical effectiveness sections of the report; Eldon Spackman, David Epstein and Stephen Palmer wrote the cost effectiveness sections of the report and performed the economic modelling. Rocio Rodriguez-Lopez wrote the sections of the report dealing with search strategies and provided information support.

Table of Contents

List of abbreviations 11

1 Summary 13

1.1 Critique of the decision problem in the manufacturer's submission 13

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer 13

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted 15

1.4 Summary of cost effectiveness submitted evidence by the manufacturer 17

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted 18

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer 19

1.6.1 Strengths 19

1.6.2 Weaknesses and areas of uncertainty 19

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG 21

2 Background 22

2.1 Critique of manufacturer's description of underlying health problem 22

2.2 Critique of manufacturer's overview of current service provision 22

3 Critique of manufacturer's definition of decision problem 23

3.1 Population 23

3.2 Intervention 23

3.3 Comparators 24

3.4 Outcomes 24

3.5 Other relevant factors 25

4 Clinical Effectiveness 26

4.1 Critique of the methods of review(s) 26

4.1.1 Searches 26

4.1.2 Inclusion criteria 27

4.1.3 Data extraction 27

4.1.4 Quality assessment 27

4.1.5 Evidence synthesis 28

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these) 28

4.2.1 Overview of the trials in the submission 28

4.2.2 Effectiveness: Placebo controlled trials 29

4.2.2.1 Placebo controlled trials study design 30

4.2.2.2 Placebo controlled trials population 30

4.2.2.3 Placebo controlled trials outcomes 32

4.2.2.4 Placebo controlled trials internal validity 33

4.2.2.5 Meta-analysis of placebo-controlled trials 34

4.2.2.6 Subgroup analyses 35
4.2.2.7 Extension Studies 35
4.2.2.8 Effectiveness results: placebo-controlled trials 35
4.2.3 Effectiveness: Active Comparison 40
4.2.3.1 Study design 40
4.2.3.2 Study population 40
4.2.3.3 Study outcomes 41
4.2.3.4 Study internal validity 41
[redacted] 42
4.2.3.5 Effectiveness Results: Active comparator trial 42
4.2.4 Adverse effects of teriflunomide: all trials 45
4.2.4.1 Long-term safety of leflunomide 47
4.3 Additional trials 47
4.3.1 Searches and trial selection 48
4.3.2 Missing trials 49
4.3.3 Included trials 51
4.4 Critique of the indirect comparison and/or multiple treatment comparison 51
4.4.1 Methods 51
4.4.1.1 Heterogeneity 52
4.4.1.2 Consistency 53
4.4.2 Results of MTC 54
4.4.2.1 Discontinuations 58
4.4.2.2 [redacted] 58
4.4.3 Main conclusions of critique of MTC 60
4.5 Additional work on clinical effectiveness undertaken by the ERG 60
4.6 Conclusions of the clinical effectiveness section 60
5 Cost Effectiveness 62
5.1 ERG comment on manufacturer's review of cost-effectiveness evidence 63
5.1.1 Searches 63
5.1.2 Conclusions of the cost effectiveness review 64
5.2 ERG's summary and critique of manufacturer's submitted economic evaluation 64
5.2.1 The manufacturer's economic evaluation compared with the NICE reference case checklist 65
5.2.2 Population 67
5.2.3 Interventions and comparators 69
5.2.4 Perspective, time horizon and discounting 70
5.2.5 Model structure 71
5.2.6 Natural History 73

5.2.6.1 Disability Progression 74
5.2.6.2 Conversion from RRMS to SPMS 78
5.2.6.3 Relapse rates and severity 80
5.2.6.4 Mortality 81
5.2.7 Treatment effectiveness and extrapolation 82
5.2.7.1 Disability Progression 82
5.2.7.2 Conversion from RRMS to SPMS 84
5.2.7.3 Relapse rate and severity 84
5.2.7.4 Discontinuation 86
5.2.7.5 Mortality 87
5.2.7.6 Adverse events 88
5.2.7.7 General issues concerning treatment effectiveness 88
5.2.8 Health related quality of life 89
5.2.8.1 HRQoL of Disability States 89
5.2.8.2 HRQoL of Relapse 91
5.2.8.3 HRQoL of Adverse Events 91
5.2.8.4 HRQoL of Carers 93
5.2.9 Resources and costs 93
5.2.9.1 Disease costs 93
5.2.9.2 Drug acquisition costs 95
5.2.9.3 Administration costs 95
5.2.9.4 Monitoring costs 96
5.2.9.5 Adverse event costs 96
5.2.10 Cost effectiveness results 97
5.2.11 Sensitivity analysis 99
5.2.11.1 Patient access scheme 101
5.2.12 Subgroup analysis 102
5.2.12.1 Patient access scheme applied to subgroups 103
5.2.13 Model validation and face validity check 103
5.3 Exploratory and sensitivity analyses undertaken by the ERG 104
5.3.1 External validation 105
5.3.2 Exploration of QALYs 107
5.3.2.1 Health-related quality-of-life 107
5.3.2.2 Initial patient distribution 107
5.3.2.3 Natural history of progression 107
5.3.2.4 Conclusions of the exploration of QALYs 108
5.3.3 Further exploratory analyses 110

5.4 Conclusions of the cost effectiveness section 115
6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG 115
6.1 Overview 115
6.2 Additional ERG analyses 116
6.2.1 Population 116
6.2.2 Comparator 117
6.2.3 Natural history 118
6.2.3.1 Disability progression 118
6.2.3.2 Conversion to SPMS 120
6.2.4 Treatment effect 123
6.2.5 Costs 124
6.2.6 Health-related quality of life 125
6.2.7 Sub-group analysis 127
6.2.8 ERG's preferred analysis 128
6.3 Conclusions from ERG analyses 132
7 End of life 134
8 Overall conclusions 134
8.1 Implications for research 134
9 References 135
10 Appendices 138

Table of Tables

Table 1: Overview of teriflunomide trials 29
Table 2: Placebo-controlled trials and RSS populations 31
Table 3: Quality assessment placebo-controlled trials 33
Table 4: Effectiveness results 36
Table 5: Improvement after 3 month disability progression in TEMSO and TOWER 38
Table 6: HA and RES subgroup results 39
Table 7: Population of TENERE compared to placebo-controlled trials and RSS 40
Table 8: Quality assessment of TENERE trial 41
Table 9: Results of TENERE trial 43
Table 10: Analysis of time to failure 44
Table 11: Adverse events in placebo trials 46
Table 12: Adverse events in TENERE 46
Table 13: Breakdown of adverse events in TENERE 47
Table 14 Trials excluded from the base case but included in the 'All Years' analysis 49
Table 15: Range of potential confounding factors across the trials in the MTC (base case) 53
Table 16: Range of potential confounding factors across the trials in the MTC (All Years analysis)) 53
Table 17: MTC results ARR 55
Table 18: MTC results 3mth SAD 56
Table 19: Table MTC results Discontinuations 58
Table 20: ERG calculated 3 month SAD and discontinuations 59
Table 21: Summary of the Manufacturer's economic evaluation (and signposts to manufacturer's submission) 64
Table 22: NICE reference case checklist 66
Table 23: Patient characteristics across studies 68
Table 24: Structural assumptions of the model 73
Table 25: Number of transitions per group (Manufacturer's response, Table B.7) 75
Table 26: Patients experiencing a regression after their first SAD (Manufacturer's response, Table B1) 75
Table 27: Transitions by EDSS state for 'all RRMS' patients from the London Ontario data. 76
Table 28: Transitions by EDSS state for 'active RRMS' patients from the London Ontario data. 76
Table 29: Transitions by EDSS state from the TEMSO & TOWER placebo arms. 77
Table 30: Transition to SPMS by EDSS state 79
Table 31: Natural history relapse rates by EDSS score and MS classification 81
Table 32: Mortality multipliers by EDSS state 82
Table 33: Hazard ratios of progression 83
Table 34: Relative rate of relapse 85

Table 35: Relative risk of relapse leading to hospitalisation 86
Table 36: Annual all-cause withdrawal 87
Table 37: Utilities by EDSS state used in the base case model 89
Table 38: Utilities and durations of adverse events 92
Table 39: Carer disutility used in the model 93
Table 40: Costs associated with relapse in the economic model 94
Table 41: Acquisition costs (Manufacturer's submission, Table B7.5.1) 95
Table 42: Administration costs (Manufacturer's submission, Table B7.5.2) 96
Table 43: Cost of treatment adverse events 97
Table 44: Deterministic results comparing the manufacturer's original base case to the corrected base case 97
Table 45: Probabilistic results comparing the manufacturer's original base case to the corrected base case 98
Table 46: Deterministic results comparing the manufacturer's original base case to the corrected base case in a fully incremental analysis 98
Table 47: Probabilistic results comparing the manufacturer's original base case to the corrected base case in a fully incremental analysis 99
Table 48: Base case probabilistic cost-effectiveness results with and without the patient access scheme (corrected model) 101
Table 49: Summary of teriflunomide base case model results compared with clinical data (Manufacturer's submission, Table B.7.7.1) 103
Table 50: Comparison of clinical trial and model results (Manufacturer's response, Table B10) 104
Table 51: Base case results of NICE appraisals for RRMS 106
Table 52: HRQoL by EDSS Data from UK EQ-5D Studies 109
Table 53: A description of the exploratory analyses undertaken by the ERG 110
Table 54: Additional sensitivity analyses by ERG 113
Table 55: The results comparing differences in the initial distribution of EDSS 117
Table 56: Calculation of the blended comparator from the manufacturer's deterministic results 118
Table 57: Deterministic results using the ERG's calculation of the blended comparator from Table 56 118
Table 58: Results comparing disability progression in RRMS from London Ontario data (base case) to the BSC arms of the trials 119
Table 59: Results comparing disability progression in SPMS from London Ontario data to the BSC arms of the trials 120
Table 60: Estimates of conversion to SPMS by EDSS state 121
Table 61: Results comparing conversion to SPMS as calculated by the manufacturer to those calculated by the ERG both using the London Ontario data 123
Table 62: Results of comparing the relative treatment effect on progression of teriflunomide 124
Table 63: Results comparing the exclusion of non-health costs 125
Table 64: HRQoL by EDSS Data from UK EQ-5D Studies 126

Table 65: Results of the ERG's HRQoL scenarios 127

Table 66: Deterministic analysis of the manufacturer's corrected base case using the RRMS results of teriflunomide and the subgroup results of natalizumab and fingolimod 128

Table 67: Deterministic analysis of the manufacturer's corrected base case using the RRMS results of teriflunomide and the subgroup results of natalizumab and fingolimod using the PAS price 128

Table 68: Probabilistic results of the full incremental analysis comparing the ERG's preferred analysis, non-PAS price 129

Table 69: Probabilistic results comparing the ERG's preferred analysis using selected comparators, non-PAS price 129

Table 70: Probabilistic results of the full incremental analysis comparing the ERG's preferred analysis with the PAS price 130

Table 71: Probabilistic results comparing the ERG's preferred analysis using selected comparators, with the PAS price 130

Table 72: Deterministic results comparing the ERG's preferred analysis with the non-PAS price 131

Table 73: Deterministic results comparing the ERG's preferred analysis using selected comparators, with the non-PAS price 131

Table 74: Deterministic results comparing the ERG's preferred analysis with the PAS price 131

Table 75: Deterministic results comparing the ERG's preferred analysis using selected comparators, with the PAS price 132

Table 76: Deterministic results comparing the ERG's preferred analysis with the PAS price but using the 'post-2000' MTC. 132

Table of Figures

Figure 1: TEMSO 6 month SAD 37

[redacted] Change from baseline in EDSS over time 38

[redacted] 44

[redacted] 45

Figure 5: EDSS state distributions across studies 69

Figure 6: Model structure (Figure B7.2.1, of the manufacturer's submission)* 72

Figure 7: Average number of years in health states (ERG analysis based on manufacturer model) 77

Figure 8: Conversion of BSC cohort to SPMS 79

Figure 9: Treatment effect on SPMS population over time (ERG analysis based on the manufacturer model) 84

Figure 10: Cumulative relapse (ERG analysis based on manufacturer model) 85

Figure 11: Incremental mortality (ERG analysis based on the manufacturer model) 88

[redacted]: Comparison of HRQoL between Orme et al. and TEMSO (Manufacturer's submission, Figure B7.4.5) 90

[redacted]The ERG explores further HRQoL data and methods for combining this data in the model in Section 6.2.6. 91

Figure 14: The natural history costs of MS compared (Manufacturer's submission, Figure B7.5.4) 94

Figure 15: Probabilistic scatter plot of teriflunomide versus the blended comparator (Manufacturer's submission, Figure B7.7.3) 100

Figure 16: Multi-CEAC of teriflunomide versus all beta-interferons and glatiramer acetate (Manufacturer's submission, Figure B7.7.6) 100

Figure 17: The proportion of patients in SPMS (ERG analysis based on the manufacturer model) 122

List of abbreviations

AE	adverse event
ALT	alanine aminotransferase
ARR	annualised relapse rate
BSC	best supportive care
CEAC	cost-effectiveness acceptability curve
CHMP	Committee for Medicinal Products for Human Use
CI	confidence interval
CIS	clinically isolated syndrome
DIC	deviance information criterion
DMT	disease modifying therapy
EDSS	expanded disability status scale
EMA	European Medicines Agency
EQ-5D	European Quality of Life-5 Dimensions
ERG	evidence review group
HA	highly active
HR	hazard ratio
HRQL	health related quality of life
ICER	incremental cost-effectiveness ratio
ITT	intention to treat
MRI	magnetic resonance imaging
MS	multiple sclerosis
MTC	mixed treatment comparison
NA	not applicable
PPMS	primary progressive multiple sclerosis
PRMS	progressive relapsing multiple sclerosis
NHS	National Health Service
PSA	probabilistic sensitivity analysis

PSS	personal social services
QALY	quality adjusted life year
RCT	randomised controlled trial
RES	rapidly evolving severe
RR	relative risk
RRMS	relapsing remitting multiple sclerosis
RSS	risk sharing scheme
SPMS	secondary progressive multiple sclerosis

Drug names

To avoid confusion between the various beta-interferon drugs, in this report we have used brand names as follows:

Generic name	***Brand names***
Interferon beta-1a 30µg	Avonex 30µg 30µg
Interferon beta-1b 250µg	Betaferon 250µg
Interferon beta-1b 500µg	Betaferon 500µg
Interferon beta-1a 22µg	Rebif 22µg
Interferon beta-1a 44µg	Rebif 44µg

Sustained accumulated disability (SAD)

Throughout the document we refer to 3 month SAD rather than 12 week and to 6 month rather than 24 week for consistency.

1 Summary

1.1 Critique of the decision problem in the manufacturer's submission

The intervention as outlined in the NICE scope is teriflunomide (®Aubagio), a new, oral immunomodulatory disease modifying therapy (DMT) for the treatment of relapsing remitting multiple sclerosis (RRMS). The scope issued by NICE included RRMS, secondary progressive multiple sclerosis (SPMS) with relapses and progressive relapsing multiple sclerosis (PRMS), but as teriflunomide has a Committee for Medicinal Products for Human Use (CHMP) positive opinion for RRMS only, the focus on this group of patients is appropriate.

The manufacturer included an analysis of the subtypes of RRMS: highly active (HA) and rapidly evolving (RES) as specified in the NICE scope. However the manufacturer does not consider these the target populations for teriflunomide and anticipates that these patients will continue to receive fingolimod or natalizumab as appropriate. Teriflunomide is presented as an alternative to glatiramer acetate and to the beta-interferons as outlined in the NICE scope. The manufacturer's position appears to be reasonable in terms of the comparators given that fingolimod and natalizumab are licensed and are approved by NICE only for HA and RES. However, teriflunomide, the beta-interferons and glatiramer acetate are all licensed for a general RRMS population and not just the non-HA/RES one.

The manufacturer included all outcomes specified by NICE with the addition of resource utilisation, (including relapse rate and accumulation of disability). The two outcomes, severity of relapse and freedom from disease, were included in the submission but had not been directly assessed as endpoints in the trials. The manufacturer submitted proxy outcomes for these and acknowledges their limitations in addressing the decision problem.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

The manufacturer conducted a systematic review to identify randomised controlled trials (RCTs) involving any treatment for RRMS / PMS in adults. Three placebo-controlled trials of teriflunomide were identified (TEMSO, TOWER and Study 2001) which were pooled in meta-analyses to produce summary results. Across the placebo-controlled trials, 788 patients received 14mg teriflunomide and 813 received placebo. Two extension studies to TEMSO and Study 2001 (without a placebo-controlled arm) provided longer term safety and efficacy data. One trial, TENERE, was identified that directly compared teriflunomide to another DMT (Rebif). To supplement the direct evidence, the manufacturer undertook a mixed treatment comparison (MTC) to assess the effectiveness of teriflunomide compared to other DMTs. The results of the MTC informed the economic model assessing the cost-effectiveness of teriflunomide.

Effectiveness

The results of the meta-analysis of placebo-controlled trials showed teriflunomide significantly reduced adjusted annualised relapse rate (ARR) compared to placebo: Relative Risk (RR) = 0.664 (95% CI: 0.589 to 0.749). Patients were less likely to accumulate a clinically significant degree of disability that was sustained for at least 3 months (3 months SAD) (as a surrogate for permanent progression of disability): RR = 0.694(0.544 to 0.886). Teriflunomide did not significantly reduce 6 month SAD in the two main trials, TEMSO HR = 0.749 (95% CI: 0.505 to 1.111) and TOWER HR = 0.843 (95% CI: 0.533 to 1.334). The placebo controlled trials failed to demonstrate significant benefits of teriflunomide in terms of EDSS, Fatigue Impact Scale (FIS), and Health-related quality of life(HrQoL) (SF-36 or EQ-5D).

[redacted] However, these results cannot be taken as reliable due to [redacted] In clinical practice these patients are likely to be treated with fingolimod and natalizumab respectively.

In the only active comparator trial (TENERE) there were no significant differences between teriflunomide and Rebif 44 for the primary outcome of time to failure (confirmed relapse or permanent treatment discontinuation): HR = 0.86 (95% CI: 0.564 to 1.314). There were no significant differences in ARR: RR = 1.197 (95% CI: 0.623 to 2.299). [redacted] The MTC results for ARR indicate statistically significant treatment benefit for all DMTs compared with placebo. When compared with the other DMTs the ARR for teriflunomide was similar to Betaferon 250µg, slightly better than Avonex 30µg (but not statistically significantly so), slightly worse than both Rebif 44µg and glatiramer acetate (but again not statistically significantly so) and significantly less effective than fingolimod and even more so than natalizumab. In the MTC, compared with placebo most DMTs have a small, not statistically significant benefit on 3mth SAD, with only fingolimod and natalizumab achieving statistical significance. When compared with the other DMTs teriflunomide appeared more effective for 3 month SAD except for the base case estimate compared with natalizumab, although no result is statistically significant. Teriflunomide was [redacted] Results versus Rebif and glatiramer acetate were not consistent across the base case and All Years MTC networks. Also the effect size teriflunomide compared with placebo for 3 month SAD reported

in the MTC was greater (and more precise) than the comparisons between all the individual beta-interferons/glatiramer acetate and placebo.[redacted]

Safety

In the meta-analysis of placebo-controlled trials, patients on teriflunomide were more likely to discontinue due to adverse events: RR = [redacted]. In a pooled analysis of Study 2001 and TEMSO the most common treatment emergent adverse events (TEAE) were liver function analyses, diarrhoea, nausea and vomiting and alopecia (full list provided in the submission). Extension studies found that hepatic disorders and anaphylactic reactions were the most common class of adverse events.

In the active comparator trial, TENERE, adverse events occurred at a similar rate in both treatment groups but patients were more likely to discontinue due to adverse events with Rebif 44µg (10.9% vs. 21.8%). Among adverse events more common with teriflunomide than Rebif 44µg were diarrhoea and alopecia. Among those more common with Rebif 44µg were influenza-like illness and increase in alanine aminotransferase.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

Direct evidence of the clinical effectiveness of teriflunomide is derived from three placebo-controlled trials, meta-analyses of these trials and two open-label (without a placebo arm) extension studies. The two key placebo studies are TEMSO and TOWER, both of which are well-conducted, international trials with over 1000 participants.

There is limited direct evidence comparing teriflunomide with other DMTs. There is one relevant head to head trial, TENERE, which compared Rebif 44µg and 14mg teriflunomide. This trial is smaller than the placebo trials (n = 324), was not double-blind, and may not be powered to detect outcomes other than the primary outcome of time to failure, i.e. not powered for ARR or SAD.

The population in the trials broadly reflects the UK RRMS population and the UK Risk Sharing Scheme (RSS) but baseline mean relapse rates and EDSS scores were higher in the RSS. Gender breakdown in all trials reflected the female-dominated nature of MS. The trials primarily reflects a white population.

MS is a lifelong disease and progression of disease can occur over many years. The main trials approximated two years' duration but the longest placebo-controlled trial has *variable* follow-up to 154 weeks; the EMA recommends that trials follow all patients up for *at least* two years to fully capture the outcome of long-term disability. Equally, the relatively short-term duration of the trials may not be adequate to capture differences in relapse rate given that relapses occur relatively infrequently in MS. The phase II study included in the meta-analysis of ARR had a follow up of just

six months. The extension studies, which present up to five years follow up, suggest a continuing effect on relapse, but these studies are not placebo-controlled and thus take no account of natural history.

EDSS is used to assess SAD as outlined in the scope. However the main time point used for SAD is three months rather than six months as recommended by the EMA and which allows for patient improvement to a better health state. No meta-analysis was conducted of six month SAD for the placebo comparisons and although data from the two key trials suggest there may be some benefit, results are not statistically significant.[redacted]

Overall the treatment benefit of teriflunomide in terms of ARR and 3 month SAD was not reflected in the 6 months SAD results[redacted]Although teriflunomide showed superiority of over placebo in terms of severity of relapse and freedom of disease activity, these were not pre-specified endpoints of the trials and used surrogate outcomes.

Given the lack of direct evidence against Avonex 30µg and Betaferon 250µg and other DMTs, the manufacturer's decision to conduct a MTC was appropriate. The MTC analysis was well conducted in that it included all relevant comparators of teriflunomide, though not all comparisons are actually of direct relevance to the use of teriflunomide in the NHS at the present time, due to available marketing authorisations and restricted NICE guidance. The methods of the MTC were generally appropriate. The ERG has one major criticism which is of the decision to exclude pre-2000 trials from the base case analysis. In the opinion of the ERG some covariate analyses using the All Years network to control for potential confounders would have been more appropriate and would have avoided the loss of the placebo comparisons for the beta-interferons. Informal checks by the ERG did not identify major problems with consistency, except with for the impact on the comparison of Betaferon with placebo and to a lesser extent the comparison of teriflunomide with Rebif 44µg, generated by the smaller base case network for 3 month SAD. This inconsistency may have contributed to the favourable results for teriflunomide generated by the MTC (particularly the base case analysis). Because the relative effect of teriflunomide in terms of 3 month SAD is such a key driver in the model the MTC result for these are very important. The inconsistency was less marked with the All Years network and therefore the ERG considers this more complete network is to be preferred.

Details of adverse events in the trials were provided. However the submission did not include a synthesis of adverse effects data that could be readily checked against supporting tables. The adverse

effects data was limited as would be expected for a newly licensed drug. The submission did also make reference to the safety profile of leflunomide, the active metabolite of teriflunomide, which has been authorised for the treatment of rheumatoid arthritis and psoriatic arthritis for over a decade.

1.4 Summary of cost effectiveness submitted evidence by the manufacturer

The manufacturer undertook a systematic review of cost-effectiveness studies including teriflunomide for the treatment of patients with RRMS. No cost-effectiveness studies of teriflunomide were found in the published literature, thus a *de novo* economic model was developed. The model evaluates the cost-effectiveness of teriflunomide compared with a 'blended comparator' in the base case analysis. The blended comparator is a weighted average, by market share, of commonly used beta-interferons and glatiramer acetate currently used in this population. The manufacturer also presented results based on a fully incremental analysis comparing teriflunomide with each individual DMT. The population considered in the main analyses is patients with RRMS in accordance with the marketing authorisation. Further analyses were undertaken in the subgroup of patients with highly active disease (HA) or rapidly evolving severe disease (RES).

The cost-effectiveness evaluation uses a decision model, designed as a Markov model, to model disease progression using 20 health states representing different levels of disease severity (by tracking EDSS scores whilst in RRMS and after conversion to SPMS) and death. In the base case disability progression and conversion to SPMS were assumed irreversible. The model also accounts for the frequency and severity of relapses, adverse events, withdrawal from treatment, and death. Although the occurrence of relapse does not influence the way in which progression is modelled, relapse is modelled to be dependent on EDSS score. After withdrawing from first-line treatment patients are assumed to receive best standard care (BSC) in the base case, however, the model is flexible to test the effect of different treatments in second and third-line care.

The stated perspective of the analysis of costs is that of the NHS and PSS. Costs are separated into disease costs, administration and monitoring costs, drug acquisition costs and adverse event costs. QALYs are used as the measure of outcomes. HRQoL varies by disease severity and accounts for both patients and carers utility. HRQoL adjustments are also applied to account for relapses and adverse events. Patients are modelled to discontinue treatment because of disease progression to an EDSS score of 7 or greater, conversion to SPMS or death. Patients also stop treatment in the model according to trial withdrawal rates. A 50 year time horizon is used in the model; both costs and benefits are discounted at 3.5%.

Natural history data are derived from an external observational data set and are used to inform the key events in the model. Natural history is assumed to represent the course of disease under BSC. Treatment effects are derived from a mixed treatment comparison and applied as hazard ratios to the

external observational data set. Treatment effects are assumed to be sustained for as long as patients remain on treatment. Data from the placebo arm of the clinical trials (TEMSO and TOWER) is also available in the model to inform the natural history of disease and is tested in sensitivity analyses.

Data on mortality are derived from national mortality statistics and are adjusted for the additional risk of mortality for different EDSS states. The adjustment factors were derived from a published study (Pokorski ,1997).[1]

Both deterministic and probabilistic sensitivity analysis were carried out by the manufacturer to demonstrate the level of uncertainty around the model results. The incremental cost-effectiveness ratio (ICER) of teriflunomide relative to the 'blended comparator' is estimated to be [redacted] per QALY gained in the deterministic analysis and [redacted] per QALY gained in the probabilistic analysis. The ICERs for teriflunomide in the full incremental analysis are reported by the manufacturer as [redacted] and [redacted] per QALY gained compared to Rebif 22µg for the deterministic and probabilistic analyses respectively. In the model the deterministic result is slightly different than that reported by the manufacturer because of rounding and is actually £[redacted] per QALY gained compared to Rebif 22µg.

In response to the ERG's request for further evidence, the manufacturer submitted a new version of the Excel-based economic evaluation having corrected for an error in a calculation. The corrected ICER of teriflunomide relative to the 'blended comparator' was estimated to be £[redacted] per QALY gained in the deterministic analysis and £[redacted] per QALY gained in the probabilistic analysis. The ICERs for teriflunomide in the full incremental analysis were £[redacted] for the deterministic analysis; in the probabilistic analysis teriflunomide was dominated (had lower QALYs and higher costs) than Rebif 22µg.

The manufacturer also submitted a price discount patient access scheme (PAS). The proposed list price for teriflunomide 14mg is £13,529 per patient per year, however the manufacturer is offering a fixed price to NHS patients of [redacted] per patient per year. Using the PAS price the manufacturer demonstrates that teriflunomide dominates its comparators (i.e. it is less costly and more effective).

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG critiques a number of issues which challenge the validity of the manufacturer's analysis.

The ERG found that the data used to populate the natural history and baseline characteristics used in the model represented a more severe population than is expected to be treated with first-line treatment. The manufacturer includes non-health costs in the model with very little detail provided on what is included in these costs. The ERG considered this to be contrary to the NICE reference case.

The main comparator used in the model is a blended comparator. The ERG considers the blended comparator inappropriate as it does not meet the NICE reference case that requires the use of a fully incremental analysis.

The manufacturer has positioned teriflunomide as a first-line treatment to be used for patients that would otherwise take beta-interferons or glatiramer acetates. They further state that fingolimod and natalizumab are not comparators of teriflunomide because these treatments have been approved to treat HA or RES patients. This suggests that teriflunomide will be used in non-HA/RES RRMS patients. However, no evidence is provided on this patient population.

The main source for the treatment effects used in this model is the manufacturer's post-2000 MTC. This MTC excluded trials pre-2000 due to the difference in diagnostic criteria used in MS clinical trials and the concern that pre-2000 trials may have included patients more severe than post-2000 patients. The choice of trials included in the MTC are not well justified and further analysis of this issue could provide useful information.

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer

1.6.1 Strengths

The manufacturer presented a well-conducted systematic review of RCTs involving any treatment for RRMS / PMS in adults. Three placebo-controlled trials were identified for the review (TEMSO, TOWER and Study 2001) which were pooled in meta-analyses to produce summary results. The two key placebo studies are TEMSO and TOWER, both of which are well-conducted, international trials with over 1000 participants. These were supplemented by two extension studies to TEMSO and Study 2001 (without a placebo-controlled arm) which provided longer term safety and efficacy data. The submission also included a trial that directly compared teriflunomide to another DMT (Rebif). To supplement the direct evidence, the manufacturer undertook a MTC to assess the effectiveness of teriflunomide compared to other DMTs. The results of the MTC informed the economic model assessing the cost-effectiveness of teriflunomide.

The manufacturer undertook a comprehensive well-conducted systematic review of the cost-effectiveness of teriflunomide. A *de novo* cost-effectiveness model was presented that was flexible to sensitivity and scenario analyses. Multiple analyses were considered and presented by the manufacturer. In response to the points for clarification the manufacturer presented additional evidence that was useful for informing the analyses.

1.6.2 Weaknesses and areas of uncertainty

Although the placebo-controlled trials were well-conducted they were of short duration given the lifelong duration of MS and the infrequency of relapses. The active comparator trial was smaller and

underpowered for the main outcomes of interest to this submission (ARR and SAD) and it was not double-blinded and therefore the trial could have been biased in favour of teriflunomide.

There is still uncertainty on the ability of teriflunomide to delay or reduce accumulation of disability. The outcome measure for accumulation of disability is SAD for three months rather than six months recommended by the EMA: 3 month SAD is not a adequate reflection of permanent accumulation of disability. Also the treatment benefit of teriflunomide in terms of ARR and 3 month SAD was not reflected in the 6 months SAD results, [redacted] none of which demonstrated benefit compared with placebo.

Although severity of relapse and freedom of disease activity showed superiority of teriflunomide over placebo, these were not pre-specified endpoints of the trials, used surrogate outcomes and no meta-analysis was undertaken.

There is a lack of reliable direct evidence comparing teriflunomide with the beta-interferons and glatiramer acetate particularly in terms of a robust measure of disability accumulation such as 6 month SAD.

The submission did not include a synthesis of adverse effects data that could be readily checked against supporting tables. The relatively short duration of the placebo-controlled trials impacts on assessment of any differences in mortality and rarer adverse events.

In summary, the clinical evidence suggests that teriflunomide appears to have some benefits compared to placebo, notably for short-term relapse rate but its role in preventing accumulation of disability is less clear. There is a lack of reliable direct evidence comparing teriflunomide with other DMTs. Despite the existence of a comparative trial, the relative effectiveness and tolerability of teriflunomide and Rebif 44µg is uncertain.

The population represented in the manufacturer's base case economic analysis is considered by the ERG to be more severe than those who are likely to use teriflunomide. The ERG are concerned with the initial distribution of EDSS used in the base case analysis, the natural history progression rates and the proportion HA and RES patients included.

In the model patients are allowed to start teriflunomide at EDSS 6, however, the ERG consider the initial distribution of EDSS will be more similar to those who were entered into the trial. Similarly, the main source of natural history data used in this model has been previously criticised and does not seem to represent the patients eligible for treatment with teriflunomide. Of particular concern is the manufacturer's exclusion of improvements in disability progression (i.e. regression in the EDSS

matrix) in the base case analysis. The population of interest to the manufacturer is those patients that would otherwise be treated with glatiramer acetate of beta-interferons, but not natalizumab or fingolimod. As natalizumab and fingolimod are approved in the RES and HA populations the population of interest in this analysis is the RRMS non-HA/RES population, however no evidence on this population has been presented.

No exploration of SPMS transitions was undertaken by the manufacturer. A single source of data was used to calculate SPMS transitions that did not allow for disability to improve, no sensitivity analysis was undertaken by the manufacturer around this model input.

The cost-effectiveness results highly depend on the trials chosen in the MTC. Further analysis controlling for the trials heterogeneity could have been taken into account rather than excluding pre-2000 trials.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG's preferred analysis results in teriflunomide being dominated (i.e. more costly and less effective) compared to glatiramer acetate using the non-PAS price and an ICER of £[redacted] per QALY gained compared to glatiramer acetate using the PAS price.

The ERG's preferred analysis combined population characteristics, HRQoL values and natural history data from the TEMSO trial that the ERG considers to be more reflective of the first-line patients eligible for teriflunomide. The ERG recalculated the SPMS conversion and excluded non-health costs and chose to use the MTC results which included all trials, pre and post-2000.

Whether or not to include the pre-2000 trials in the MTC was the most influential parameter on the conclusions. Combining the ERG's preferred analysis except using the manufacturer's preferred MTC which excludes pre-2000 trials results in an ICER of £[redacted] per QALY gained using the PAS price.

The ERG also found that the natural history of progression in SPMS was influential on the model results; however, no additional data was available for the ERG to use in a preferred analysis. If SPMS patients are expected to transition more slowly or improve their disability state (EDSS score) than the manufacturer's and ERG's estimates of the ICER may be underestimated.

The manufacturer states that teriflunomide will not be used in HA and RES populations where fingolimod and natalizumab are more appropriate, however no evidence is provided for the non-HA/RES RRMS population. Given the high proportion of patients with HA or RES in the TOWER trial and the subgroup analyses which suggest teriflunomide is more effective in these populations, the effectiveness of teriflunomide in the non-HA/RES RRMS population may be overestimated in this

analysis resulting in an underestimation of the ICER. However the applicability of the TOWER HA and RES subgroups are unclear given the difference in the definitions of these populations used in this trial.

2 Background

2.1 Critique of manufacturer's description of underlying health problem

The manufacturer provided an accurate and appropriately referenced overview of MS and its classifications into the three sub-types: relapsing-remitting (RRMS), primary progressive (PPMS) and secondary progressive (SPMS). It is noted that at the date of this report the manufacturer has a CHMP positive opinion for RRMS only and the submission focuses on this group of patients. The manufacturer accurately described the specific problems experienced by patients with RRMS.

The role of the Expanded Disability Status Scale (EDSS) in the measurement of the accumulation of disability in RRMS was highlighted. This is a 10 point instrument that measures different areas of functional disability from 0 to 5 and focuses on ambulatory disability from 6 to 10. EDSS scores are used in the trials in this submission for the assessment of relapses and accumulated disability and are key to measuring effectiveness and underpin the economic model. See section 4.2 for further details.

2.2 Critique of manufacturer's overview of current service provision

The current treatment options for modification of disease in MS are appropriately outlined by the manufacturer and make reference to all the comparators in the decision problem. It is noted, however, that the manufacturer considers teriflunomide as an alternative option only to glatiramer acetate and beta-interferon which are available under the Department of Health Risk Sharing Scheme.[2] Glatiramer acetate and beta-interferon are licensed for the whole RRMS population but are only available as injectables. The manufacturer fairly describes some of the limitations of these injectables including inconvenient mode of administration and issues of tolerability and, in some cases, lack of response to treatment.

Subgroup analyses of patients with highly active MS and rapidly evolving MS are presented in the submission where teriflunomide is compared to fingolimod and natalizumab which are both licensed only for these two subtypes of RRMS and these drugs form part of the manufacturer's MTC. However the manufacturer does not consider the teriflunomide to be appropriate as an alternative option to these drugs. The treatment pathway outlined by the manufacturers indicates that it is expected that teriflunomide will be considered in practice as an alternative treatment option to interferon-beta or glatiramer acetate for those with a diagnosis of RRMS that is not RES / HA disease. However, teriflunomide, the beta-interferons and glatiramer acetate are all licensed for a general RRMS population and not just the non-HA/RES one.

The manufacturer estimates that 31% of those with RRMS in the UK are treated with a DMT[3] The manufacturer's advisory board further suggest that 25% of patients are resistant to injectables.[4] This figure is very difficult to estimate as it depends on the definition of resistance. The clinical advisor to the ERG identified several groups of patients who might be deemed 'resistant': patients who decline treatment and hence never start treatment with a DMT, those who start treatment with an injectable but cannot continue due to needle phobia or inability to inject, those who have to stop treatment due to skin necrosis/abscess, and those who run out of injection sites over time. The number in the latter group will increase over time. We have been advised that uptake of an oral drug will potentially be high due to increased acceptability of the mode of administration. Overall numbers relating to uptake of teriflunomide are uncertain.

The manufacturer notes on-going NICE appraisals regarding RRMS for the following therapies: alemtuzumab, laquinimod and dimethyl fumarate. It is noted that these drugs are included in the manufacturer's MTC. However the MTC results for these drugs are not reported by the manufacturer and neither are these drugs included in the economic model. The ERG has conducted this assessment independently of the other ongoing NICE appraisals.

3 Critique of manufacturer's definition of decision problem

3.1 Population

The scope issued by NICE included RRMS, SPMS and PPMS but teriflunomide is licenced only for RRMS and so the restriction of the decision problem to RRMS is appropriate. Highly active MS (HA) and rapidly evolving MS (RES) are subtypes of RRMS. The manufacturer included a subgroup analysis of HA and RES because these are specified in the NICE scope. However the manufacturer does not consider these the target populations for teriflunomide as these groups will be treated with fingolimod or natalizumab. The ERG considers it reasonable that these subgroups are included and recognises the limitations of the analyses of these groups.

3.2 Intervention

The intervention, as outlined in the scope, is teriflunomide, a new, oral, immunomodulatory DMT for the treatment of RRMS. Teriflunomide (®Aubagio) is administered as a 14mg film-coated tablet once daily. On 21 March 2013, the Committee for Medicinal Products for Human Use (CHMP) adopted a positive opinion, recommending the granting of a marketing authorisation. The approved indication for teriflunomide is: "treatment of adult patients with relapsing remitting multiple sclerosis (MS)".[5] The Committee also concluded at that time that the active substance, teriflunomide, which is a metabolite of leflunomide, could not be considered a new active substance. Following a request by the manufacturer, CHMP re-examined the initial opinion and concluded on 27 June 2013 that

teriflunomide could be considered to be a new active substance in view of differences between teriflunomide and leflunomide as regards safety.

3.3 Comparators

The NICE scope specified comparisons with beta-interferon, glatiramer acetate, fingolimod and natalizumab; these comparators are all included in the submission. Subgroup analyses of patients with highly active MS and rapidly evolving MS are presented in the submission where teriflunomide is compared to fingolimod and natalizumab and these drugs form part of the manufacturer' MTC. However, the manufacturer does not consider teriflunomide to be appropriate as an alternative option to these drugs, stating that, "It is not anticipated that teriflunomide will be routinely used in these subgroups of MS" (i.e. HA and RES). The treatment pathway outlined by the manufacturer indicates that it is expected that teriflunomide will be considered in practice as an alternative treatment option to interferon-beta or glatiramer. The manufacturer's position appears to be reasonable in terms of the comparators, given that fingolimod and natalizumab are licensed and are approved by NICE only for HA and RES. However, teriflunomide, the beta-interferons and glatiramer acetate are all licensed for a general RRMS population and not just the non-HA/RES population. The clinical advisor to the ERG advised that in practice HA patients are defined as those who have failed on a beta-interferon and would go on to be treated with galtiramer acetate or fingolimod. RES patients are more easily defined being patients with 2 disabling relapses in 1-year and evidence of new or Gd-enhancing lesions on MRI and would be treated with natalizumab.

3.4 Outcomes

The manufacturer included all outcomes specified by NICE as follows with the addition of resource utilisation: relapse rate, severity of relapse, disability, symptoms of MS, freedom of disease activity, mortality, adverse effects of treatment and health-related quality of life.

Aspects of the decision problem could not be fully answered by the manufacturer's submission: severity of relapse and freedom from disease activity were not predefined endpoints of the teriflunomide trials. The manufacturer supplied a range of supplementary data to address the outcome severity of relapse, including relapses requiring hospitalisation and occurrence and risk of sequelae and ARRs requiring IV corticosteroid treatment. This appears reasonable and the manufacturer acknowledges the limitation of the surrogate outcomes and does not pool results for severity of relapse. For freedom of disease activity, the manufacturer supplied data on patients who remained relapse free and MRI results where available. The results for the two outcomes of severity of relapse and freedom of disease activity should, therefore, be treated with some caution.

Although quality of life and mortality are addressed as outcomes, the randomised evidence is limited to two years follow up which is insufficient to address these outcomes. These data are supplemented

by the longer-term extension studies but these studies do not take account of natural history of disease in relation to these outcomes.

The two key outcomes of this submission, confirmed relapse and sustained accumulation of disability (SAD), are measured by changes in EDSS scores. This is a 10 point instrument that measures different areas of functional disability from 0 to 5 and focuses on ambulatory disability from 6 to 10. The EMA highlight the limitations of EDSS for measuring disability in MS but advise that it should still be used to provide a comparator for existing studies.[6] In the trials in this submission, confirmed relapses required an increase of 1 point in each of two EDSS functional system scores or 2 points in one EDSS functional system score (excluding bowel and bladder function and cerebral function) or an increase of 0.5 points in EDSS score from previously stable assessment. If the EDSS score was previously 0, an increase in EDSS score of 1 point was required. Sustained accumulation of disability was defined as an increase from baseline of at least 1.0 point in EDSS score for patients with a baseline EDSS score of ≤ 5.5 or at least 0.5 points for patients with baseline EDSS score > 5.5 for at least 12 weeks (3 month SAD). In addition to reporting the outcomes of relapse and SAD, this report will include detail on changes in EDSS.

The submission focuses on 3 month SAD in the trials and the MTC. However 6 month SAD is preferred as an outcome measure[6] as after 3 months there is still a strong possibility of improvement to a better EDSS state.

3.5 Other relevant factors

The manufacturer submitted a price discount patient access scheme (PAS). The proposed list price for teriflunomide 14mg is £13,529 per patient per year, however the manufacturer is offering a fixed price to NHS patients of [redacted] per patient per year. This PAS has been approved by the Department of Health.

4 Clinical Effectiveness

The manufacturer conducted a systematic review to identify RCTs involving any DMT for the treatment of RRMS / PMS in adults. Trials involving teriflunomide were then identified, extracted, assessed and synthesised in a meta-analysis. These trials and all others identified for the review were combined in a mixed treatment comparison (MTC), the estimates from which were used in the economic modelling.

This section contains a critique of the methods of the review(s) of clinical effectiveness, followed by a description and critique of the teriflunomide trials included in the review, including a summary of their quality and results and the results of any synthesis of studies. A critique of the manufacturer's MTC is also included.

4.1 Critique of the methods of review(s)

4.1.1 Searches

The manufacturer's submission described the search strategies used to identify relevant clinical effectiveness studies about the use of teriflunomide for the treatment of RRMS in adults. Search strategies were briefly described in the main body of the submission; however full details were provided in the Appendices.

The electronic databases MEDLINE and MEDLINE In-Process (via Pub Med), EMBASE(via Embase.com) and Cochrane Central Register of Controlled Trials (CENTRAL) were searched to identify clinical studies on the use of teriflunomide and other comparators. In addition to this, grey literature was searched in several medical society and regulatory body websites and a hand searched was carried out.

Searches were conducted in May 2012 and subsequently updated on 12 November 2012. Search strategies for each database were documented in Tables C10.2.1-3. The searches covered the period 1st January 1980 to November 12 2012, were limited to English language publications, and excluded letters, editorials, comments and animal studies.

Overall the searches were appropriate and well documented, and included the use of both subject indexing terms (MeSH and EMTREE) and free text searching. Field searching, Boolean operators and truncation were used where required. All the databases required by NICE were searched, though only CENTRAL was searched in the Cochrane Library when it might have been useful to have searched the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE) and the Health Technology Assessment (HTA) database. For this reason the manufacturer missed a protocol of a Cochrane Systematic Review on teriflunomide for the treatment of multiple sclerosis that is finished and published now.

The trade name of teriflunomide and the trade name of its comparators were not included in the search strategies. It is possible (though not likely) that potentially useful records were not retrieved. The search strategies used in the manufacturer's submission were limited to experimental trials. However a search for other study designs such as cohort or case control studies may have provided useful supplementary information about safety. It is not clear if the methodological search filters used in PubMed, EMBASE were derived from validated search filters.

4.1.2 Inclusion criteria

The manufacturer outlined appropriate inclusion criteria for population, interventions and comparators, outcomes and study designs. It is noted that eligibility criteria for the review included patients with SPMS. All included direct trials of teriflunomide had a small percentage of patients with SPMS (at least 87% were patients with RRMS). We consider this to be overall acceptable but it is noted, in contrast, that the trials in the MTC largely had 100% RRMS; see section 4.3.

The manufacturer included open-label extension studies in addition to RCTs in order to assess long-term safety of teriflunomide. This approach is acceptable. However the efficacy data from these trials should be treated with some caution as there is no concurrent placebo group and therefore no information on natural course of disease.

The restriction to English language trials only could have led to trials being missed but in this instance we are unaware of any missing trials. The manufacturer's flow diagram (Figure B6.2.1) does not include data (no of studies etc.) included and excluded from the direct evidence. The inclusion criteria for the MTC are discussed in section 4.3.

4.1.3 Data extraction

Details of data extraction methods were not provided (such as number of reviewers involved and procedures to avoid errors and bias). The ERG could not, therefore, comment on the robustness of the data extraction methods.

4.1.4 Quality assessment

The manufacturer's quality assessment was adapted from CRD guidance[7] covering: randomisation, allocation concealment, similarity of groups at outset, blinding, differential dropout, under-reporting of outcomes and use of intention to treat analysis. This tool was appropriate. Trials of both direct evidence and those included in the base case MTC were assessed. No quality data were provided for the trials included in the sensitivity analysis of the mixed treatment comparison (trials conducted post 2000 and ≥ 80% RRMS). See section 4.3. The ERG has re-assessed the validity of the direct comparison teriflunomide trials(tables XX) and provides a further critique of these trials in section 4.2.

4.1.5 Evidence synthesis

The manufacturer undertook a meta-analysis of all the three placebo-controlled trials (Study 2001, TEMSO and TOWER) using a random-effects model. A random effects model may not have been appropriate due to the small number of studies (2 or 3 in each analysis).[8] The manufacturer checked for statistical heterogeneity and clinical diversity and did not find major issues of heterogeneity. Pooling in general appeared to be appropriate but please see section 4.2 for specific comments on the meta-analysis.

The study of teriflunomide vs. Rebif 44µg (TENERE) was not included in the meta-analysis as there was no placebo arm. This was appropriate. Results of the trial were tabulated and synthesised narratively.

To supplement the direct evidence the manufacturer conducted a MTC indirectly comparing eleven licensed and unlicensed drugs including beta-interferon and glatiramer acetate. This analysis is described and critiqued in sections 4.3.and 4.4.

Safety data from RCTs and trial extensions were presented individually. A pooled analysis was conducted using adverse event data from Study 2001 and TEMSO only. A critique of methods using for pooling adverse event data can also be found in section 4.2.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Overview of the trials in the submission

Table 1 gives an overview of the teriflunomide trials presented by the manufacturer. Further details are presented in the manufacturer's submission. All trials have a 7mg teriflunomide arm in addition to a 14mg arm but teriflunomide has a positive opinion for the 14mg dose only and results in this section focus on this comparison with placebo or active comparator as highlighted by the manufacturer.

Table 1: Overview of teriflunomide trials

Study	Population	Treatment groups	Study Duration	Primary Outcome
Phase III RCTs				
TEMSO	1088 patients with RMS (McDonald) 995 RRMS 51 SPMS 42 PPMS Age 18 to 55 years ≤ EDSS 5	365 teriflunomide 7mg/day 358 teriflunomide 14mg/day 363 placebo	108 weeks	ARR
TOWER	1169 patients with MS (McDonald) [redacted] Age 18 to 55 years ≤ EDSS 5	408 teriflunomide 7mg/day 373 teriflunomide 14mg/day 389 placebo	48 to 154 weeks Optional long-term extension treatment for 48 weeks or until teriflunomide available locally	ARR
TENERE	324 patients with RMS (McDonald) [redacted] Age ≥ 18 years ≤ EDSS 5	109 teriflunomide 7mg/day 111 14mg/day 104 Rebif 44µg given by sub-cutaneous injection 3 x a week	48 to 118 weeks Optional long-term extension treatment until teriflunomide available locally	Time to failure defined as the first occurrence of confirmed relapse or permanent discontinuation for any reason whichever came first
Phase II RCT				
Study 2001	179 patients with RMS (Poser) 157 RRMS 22 SPMS with relapses Age 18 to 65 years ≤ EDSS 6	61 teriflunomide 7mg/day 57 teriflunomide 14mg/day 61 placebo	36 weeks	Combined unique active (new and persisting) lesions per MRI scan
Long-term extension studies				
TEMSO extension	742 patients who completed the TEMSO study and chose to continue to the extension ≤ EDSS 6	381 teriflunomide 7mg/day 359 teriflunomide 14mg/day	On-going	Incidence of adverse events
Study 2001 extension	147 patients who completed the core study and entered the extension ≤ EDSS 6	81 teriflunomide 7mg/day 66 teriflunomide 14mg/day	On-going	Incidence of adverse events

4.2.2 Effectiveness: Placebo controlled trials

There are three placebo-controlled trials that inform the clinical effectiveness section of this report. Study 2001 is a phase II proof of concept trial for patients with RMS. 179 patients aged 18 to 65 were randomised to receive 7mg, 14mg or placebo for 36 weeks. The primary outcome was combined unique active (new and persisting lesions per MRI scan). Other clinical outcomes including ARR and SAD were reported. TEMSO is a multinational trial (127 centres in 21 countries) comparing teriflunomide to placebo in patients with relapsing MS. 1088 patients aged 18 to 55 years of age were

randomised to receive 7mg teriflunomide, 14 mg teriflunomide or placebo once daily for 108 weeks. TEMSO was designed as a superiority trial to evaluate effectiveness in terms of annualised relapse rate and progression of disability. The trial has been published in peer reviewed journals.[9] TOWER is also an international superiority trial designed to replicate TEMSO. 1165 participants aged 18 to 55 were randomised to receive 7mg teriflunomide, 14 mg teriflunomide or placebo once daily until a fixed common end date 48 weeks after randomisation of last patient (48 to 154 weeks) As with TEMSO, the primary endpoint of TOWER was ARR and the secondary outcome of this trial was time to disability progression. TOWER is currently unpublished. Across the three placebo-controlled studies 788 patients received teriflunomide 14mg and 813 patients received placebo.

4.2.2.1 Placebo controlled trials study design

The longest placebo-controlled trial has variable follow-up up to 154 weeks. These are supplemented by the open-label extension studies which are on-going and present five year follow up data. MS is a lifelong disease with a variable course. Relapses occur relatively infrequently so trials may not be long enough to capture this outcome. The EMA suggests *at least* two years to accurately assess relapses and progression of disability.[6] Whilst the two phase III placebo-controlled trials are just adequate in this regard, the phase II trial. which is essentially a 'proof of concept' trial lasting 36 weeks, is not: in this trial relapse rates and SAD are not primary outcomes and the trial is not sufficiently long enough or likely to be adequately powered for these outcomes. The issue of including this trial in the meta-analysis is discussed later in this section.

4.2.2.2 Placebo controlled trials population

The majority of participants across the trials have RRMS (at least 87%) for which teriflunomide has a CHMP positive opinion. However a number of participants have PPMS or SPMS. It is noted that the age range of the two phase III randomised trials is 18 to 55 so evidence for older patients is sparse, though not untypical of the population to be treated with teriflunomide. In terms of disease severity all the phase III trials required patients to score 5 or less on the EDSS scale whereas the phase II and extension studies allowed EDSS up to 6, patients with more severe disease. The phase II trial (due to its date) uses the Poser criteria to identify patients with MS whilst the remaining trials use the McDonald criteria which tend to identify patients earlier with less severe disease, which is more appropriate and generalisable to current clinical practice.

Both TEMSO and TOWER had similar inclusion and exclusion criteria relating to participants. Participants needed to have at least two clinical relapses in the previous 2 years or one in the preceding year, which is in line with the ABN treatment guidelines. TEMSO required participants to be relapse-free 60 days prior to randomisation whereas for [redacted]. Both trials specified prior use of natalizumab as an exclusion criterion. Equally, patients should not have used glatiramer acetate during

the preceding 3 months (TOWER) or 4 months (TEMSO) and users of teriflunomide in the preceding 6 months were excluded in both trials. These and other exclusion criteria resulted in both TEMSO and TOWER having fewer patients with prior use of DMTs than the Phase II trial (27%, 33% and 79% respectively). The proportions in TEMSO and TOWER reflect the manufacturer's positioning of the drug as an alternative to injectables for treatment naive patients and as an option for those who cannot tolerate other DMTs. Similarities and differences between the placebo-controlled trials are summarised in the Table 2.

The clinical advisor to the ERG has advised that the population of the placebo-controlled trials broadly reflects the population of UK MS patients. The ERG also checked the UK risk sharing scheme (RSS) population[10] and compared it to the trials. See Table 2 below.

Table 2: Placebo-controlled trials and RSS populations

	Study 2001	TEMSO	TOWER	Risk Sharing scheme[10]
n	179	1088	1169	4871
Age mean (SD)	39.5	37.9 (8.8)	37.9 (9.3)	39.3
% female	72	72.2	71.1	74.9
% white	NR	97	[redacted]	
Duration disease since start of symptoms mean (SD)	NR	Median 6.83 yrs.	8.00(6.72) median 6.25	8.5 (7.4)
Duration disease since diagnosis mean (SD)	Approx 5 years	Median 3.5 years	5.16 (5.66) Median 3.17	5.2 (5.8)
EDSS baseline mean (SD)	Approx 2.25	2.5	2.7 (1.37)	3.4 (1.70) RRMS only = 3.1 (1.5)
% EDSS 3.5	NR	NR	75.7	NR
Relapses in prior two years, mean (SD)	NR	1 (median in previous 1 yr.)	2.1 (1.2)	2.9 (1.2)
% RRMS	88	91.5	97.5	85.8
Previous treatment with DMT	79%	27.0%	[redacted]	Unclear
% HA *	NR	1.9%	13.3%	NR
%RES*	NR	6.6%	21.6%	NR

*HA and RES defined more specifically in TEMSO than in TOWER therefore the proportions across the trial may not be directly comparable

Overall, the differences between the populations in the trials and the RSS are not great and the trial populations can be considered generalisable to the UK population except that base line EDSS scores and relapse rate were lower in the trials than in the RSS.Thus the trial populations may be more reflective of a treatment naïve population to be treated with teriflunomide than was studied in the initial years of the RSS.

The trials have a similar proportion of females (between 69 and 73%) which reflects the gender bias of the MS population. MS symptoms appear at around 30 years of age when in the general population

women would be most likely to be starting or supporting a family.[11] However, as is usual, pregnant women or those wishing to conceive were not represented in the trials. The manufacturer states that teriflunomide is contraindicated for those populations and should a woman become pregnant, an accelerated elimination procedure is recommended as the drug can otherwise take up to 2 years to clear. Further monitoring of the drug in those wishing to conceive or becoming pregnant is required.

Nearly all of the patients in TEMSO are white [redacted] of the TOWER population are white. In England and Wales in 2011 86% of people described themselves as of white ethnicity.[12]

4.2.2.3 Placebo controlled trials outcomes

All trials provide both efficacy and safety outcome data. The primary outcome is ARR for the two phase III placebo-controlled trials and lesions identified on MRI in the Phase II trial. Further outcomes were assessed and will be discussed in this section.

In both TEMSO and TOWER the outcome of relapse was pre-defined in the same way. A relapse was defined as the appearance of a new clinical sign or symptom or clinical worsening of a previous sign or symptom that had been stable for at least 30 days and that persisted for a minimum of 24 hours (48 hours in Study 2001) in the absence of fever. Confirmed relapses required an increase of 1 point in each of two EDSS functional system scores or 2 points in one EDSS functional system score (excluding bowel and bladder function and cerebral function) or an increase of 0.5 points in EDSS score from previously stable assessment. If the EDSS score was previously 0, an increase in EDSS score of 1 point was required. Use of EDSS to confirm relapses is appropriate.[6] ARR was analysed as adjusted ARR. This was appropriate as there is indication from analysis that the manufacturer submitted of differences between EDSS cut-offs and different regions. There was no test for interaction with treatment effect so it is unknown if the treatment effect is different in the two EDSS categories (≤3.5 and >3.5).

Severity of relapse was not a defined endpoint in the trials. As stated earlier in the report, this outcome was interpreted primarily as relapses requiring hospitalisation. The manufacturer supplied additional data on the occurrence and risk of sequelae associated with relapse in TEMSO and TOWER and for TOWER data on ARRs requiring IV corticosteroid treatment were provided. This appears reasonable and the manufacturer acknowledged the limitation of the surrogate outcomes and did not pool results for severity of relapse. The Phase II placebo-controlled trial, Study 2001, does not have data on severity of relapse.

For TEMSO and TOWER sustained disability progression (a secondary outcome in both trials) was defined as an increase from baseline of at least 1.0 point in EDSS score for patients with a baseline EDSS score of ≤ 5.5 or at least 0.5 points for patients with baseline EDSS score > 5.5 for at least 12

weeks (3 month SAD). This outcome was not assessed in Study 2001. The EMA recommends 24 week (6 month) comparative SAD data.[6]

Although not considered a meaningful outcome for clinical effectiveness, data will be presented in this report from the submission and the responses to points for clarification from the manufacturer, as EDSS changes over time are influential in the economic model.

For freedom of disease activity, the manufacturer supplied data on patients who remained relapse-free and MRI results where available. This appears reasonable and the manufacturer acknowledged (as for severity of relapse) the limitation of the surrogate outcomes and did not pool results for freedom of disease activity.

HRQoL data were collected in the form of SF-36 and EQ-5D data. Details of these were not provided in the submission.

4.2.2.4 Placebo controlled trials internal validity

The ERG reassessed the quality of the placebo-controlled trials according to the criteria used by the manufacturer and these are presented in the table below.

Table 3: Quality assessment placebo-controlled trials

Assessment Criteria	***TEMSO***		[redacted]		***Study 2001***	
	Submission	ERG view	[redacted]	[redacted]	Submission	ERG view
Was randomisation carried out appropriately?	Yes	Yes	[redacted]	[redacted]	Yes	Yes
Was the concealment of allocation adequate?	Yes	Yes	[redacted]	[redacted]	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	[redacted]	[redacted]	Yes	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation	Yes	Yes	[redacted]	[redacted]	Yes	Yes
Were there any unexpected imbalances in drop-outs between groups	No	No	[redacted]	[redacted]	Yes*	Yes
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	[redacted]	[redacted]	No	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes	[redacted]	[redacted]	Yes	Yes

*there were more discontinuations in the teriflunomide 14mg group than in 7mg or placebo.

The ERG agrees that TEMSO, TOWER and Study 2001 are well conducted in terms of randomisation, blinding, comparability of groups and methods for dealing with study dropout. Procedures for randomisation and concealment of allocation were described so that selection bias was minimised. Randomisation was stratified according to baseline EDSS score ≤ 3.5 and >3.5 in all placebo-controlled trials and also according to trial site in TEMSO and TOWER. [redacted]. With a larger number of centres, there is a greater possibility of protocol deviation. However the manufacturer reports regular site monitoring for quality purposes in both trials. The use of patient and assessor blinding in both trials protected against detection bias such that results are likely to be reliable.

Both TEMSO and TOWER are large trials and both present power calculations for their primary outcomes (ARR in TEMSO) and (ARR and time to first sustained disability progression in TOWER). Study 2001 is small and not powered for these outcomes. Methods to ensure completeness of outcome data are reported in TEMSO and TOWER. EDSS scores were measured at baseline and every 12 weeks and at any unscheduled visits (for assessment of possible relapse). Patients were required to visit the study site within 7 days after onset of suspected relapse for assessment by examining neurologist. Methylprednisolone sodium succinate 1 g, intravenously daily for 3-5 days was the allowed treatment for relapses in TEMSO [redacted].

4.2.2.5 Meta-analysis of placebo-controlled trials

The manufacturer presented the results of meta-analyses for five outcomes: ARR, proportion of patients relapse-free, SAD 3 month, all-cause discontinuations and disontinuations due to adverse events. No meta-analysis of 6 month SAD was presented although data were available for both TOWER and TEMSO. The importance of this outcome in relation to sustained disability has already been highlighted. Meta-analyses were based on all three placebo-controlled trials: Study 2001, TEMSO and TOWER. The outcome of 3 month SAD was analysed without Study 2001

The meta-analysis was conducted using a random-effects model and employed an empirical Bayes estimator of the random effects variance rather than reporting the more usual Chi squared or I^2 values for heterogeneity. The submission reports values of tau (between study variance) which are very low for all outcomes except for the meta-analysis of discontinuations due to AEs. However, with only two or three trials in the meta-analysis there is insufficient power to calculate tau. The submission states that a manual check of demographic details of patients included in these trials did not flag up any clinical heterogeneity in terms of age, gender, EDSS at baseline, time since diagnosis or number of previous relapses, other than the fact that 75% of patients in the Study 2001 , the Phase II study, had received previous DMT as opposed to a smaller number in TEMSO and TOWER (79% vs. 27% and [redacted] However, some additional differences between Study 2001and the Phase III TEMSO and

TOWER are worthy of note. Study 2001, as a proof of concept study, was small (61 patients per treatment arm) and study duration was just six months so relapse rates may not be robust. Also EDSS scores were higher and a greater number discontinued treatment in the teriflunomide arm of this study. These differences suggest that pooling results of Study 2001 with TOWER and TEMSO is subject to clinical heterogeneity, which could result in statistical heterogeneity. Study 2001 was excluded from the 3 month SAD meta-analysis due to the limited duration of this trial, but it is questionable as to whether it should have been included in other analyses given its short duration and differences in previous treatments.

4.2.2.6 Subgroup analyses

Subgroup analyses of TEMSO data were performed for ARR and disability progression by baseline demographics, disease characteristics and prior use of MS drugs. Test for interactions were not statistically significant and no significant differences were apparent from the results provided to indicate potential effect modifiers.

Post-hoc subgroup analyses for patients with RES and HA disease performed to conform to NICE scope. The manufacturer anticipates that those with a previous diagnosis of RES or HA to be prescribed alternative DMTs (fingolimod or natalizumab as appropriate). Analyses were provided for the TEMSO sub-populations as this was thought to better match the criteria for these subtypes. In response to the point for clarification letter the manufacturer supplied the corresponding results for TOWER, but advised that the HA and RES subgroups had not been defined using the same criteria as in TEMSO, i.e. MRI data were not available in this trial to confirm subtypes. Consequently it is not clear that the same types of patients were included, which may explain the large difference in the proportion of patients in these sub-groups between the two trials. Limitations of subgroup analyses of HA / RES are acknowledged by the manufacturer.

4.2.2.7 Extension Studies

Study 2001 (the Phase II trial) and TEMSO have on-going long-term extension studies designed to assess long-term safety and tolerability. For Study 2001 extension, 147 patients who completed the core study were randomised to 7 or 14mg daily teriflunomide. There was no placebo arm. For the TEMSO extension 742 patients originally allocated to placebo were randomised to 7mg or 14mg teriflunomide whilst patients already on teriflunomide continued on their original dose. Across the two studies, 425 participants are receiving 14mg teriflunomide. Data are available for study 2001 at a median follow up of 7.1 years and for TEMSO at up to 5 years treatment follow up.

4.2.2.8 Effectiveness results: placebo-controlled trials

The results of the placebo-controlled trials, meta-analyses and extension studies are given in Table 4.

Table 4: Effectiveness results

Trial	TEMSO	[redacted]	Study 2001	Meta-analysis	Study 2001 extension	[redacted]
Relapse rate	Adjusted ARR* teriflunomide 0.37 (95% CI: 0.31 to 0.44) vs. Placebo 0.54 (95% CI: 0.47 to 0.62) RR 0.69 (95% CI: 0.58 to 0.81)	Adjusted ARR* ter 0.319 (95% CI: 0.267 to 0.381) vs. placebo RR 0.637 (95% CI: 0.512 to 0.793)	ARR 0.55(SD 1.12) vs. 0.81 (SD 1.22) RR 0.68 (95% CI: 0.36 to 1.27)	RR = 0.664 (95% CI:.0.589 to 0.749	ARR = 0.181	[redacted]
Severity of relapse	ARR leading to hospitalisation 0.407(95% CI: 0.272 to 0.609) for comparison between teriflunomide and placebo	ARR leading to hospitalisation 0.664(95% CI: 0.476 to 0.925) for comparison between teriflunomide and placebo	NR	Not done	NR	[redacted]
Proportion relapse-free vs. Placebo	60.6% vs. 49.3% (no p value)	[redacted]	77% vs. 62%, p = 0.098	[redacted]	55% were relapse-free at study reporting	[redacted]
Disability for at least 12 weeks (3 month SAD)	20.2%(95% CI: 15.6 to 24.7) vs. placebo 27.3% (95% CI: 22.3 to 32.3) p = 0.03 HR = 0.70 (95% CI: 0.51 to 0.97)	11.9% vs. 16.8% HR = 0.685(95% CI 0.467 to 1.004)	NR	RR = 0.694 (95% CI: 0.544 to 0,886)***	NR	[redacted]
Disability for at least 24 weeks (6 month SAD)	12.0% vs. 16.0% HR = 0.749 (95% CI: 0.505 to 1.111)	8.9% vs10.6% HR = 0.843 (95% CI: 0.533 to 1.334)	NR	Not done	NR	[redacted]
Change in EDSS	[redacted]	[redacted]				
Fatigue (FIS change from baseline)	3.8 (1.7) vs. 4.3 (1.7), p = 0.83	[redacted]	NR	Not done	NR	[redacted]
Discontinuation any cause	26.5% vs. NR	[redacted]	21% vs. 7%	[redacted]	45.4%	[redacted]
Quality of	No statistically	[redacted]	NR	Not done	NR	[redacted]

Trial	TEMSO	[redacted]	Study 2001	Meta-analysis	Study 2001 extension	[redacted]
life	significant results for SF-36 or EQ-5D	[redacted]				[redacted]

*Adjusted for treatment, EDSS score at baseline and geographic region; **RR given in the meta-analysis did not appear to match the proportion relapse-free in the trial; ***Reported as RR but should be HR

In both TEMSO and TOWER, teriflunomide significantly reduced adjusted ARR compared to placebo. Results were not significant in Study 2001. The meta-analysis confirmed the statistically significant results.

In regard to severity of relapses, in both TOWER and TEMSO teriflunomide led to statistically significantly fewer hospitalisations due to ARR than in the placebo group. This outcome was not pooled in a meta-analysis. Freedom of disease activity was presented as end of study proportion of patients relapse-free. In TEMSO a post-hoc analysis of 'freedom from disease activity' (a composite of relapse-free, disease progression free and MRI-activity free) reported 22.9% 'free of disease'in the teriflunomide group vs. 14.3% in the placebo group (no p values reported). [redacted] In both of the extension studies relapse rates were low (0.181 and [redacted]). A similar number to the placebo-controlled trials remained relapse-free.

3 month SAD was statistically significantly reduced in TEMSO [redacted]. Study 2001 did not assess this outcome. Meta-analysis confirmed the statistically significant reduction. However there is evidence that the benefit seen at 3 months may not be sustained over a longer period. The more robust measure of disability progression, 6 month SAD, was not statistically significantly reduced in TEMSO or in TOWER. Additionally the Kaplan-Meier plots supplied by the manufacturer showed little benefit of teriflunomide for 6 month SAD. The TEMSO plot is shown below.

[redacted]TEMSO 6 month SAD

[redacted]

Furthermore, the manufacturer supplied an additional table showing the proportion who improved (their disability progression regressed) after experiencing 3 month SAD although the manufacturer's model assumes that 3 month SAD is permanent. It can be seen from Table 6 that a large proportion in both groups did not have persistent disability.

Table 5: Improvement after 3 month disability progression in TEMSO and TOWER

[redacted]	[redacted]		[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The benefit of teriflunamide on terms of preventing disability progression is also not reflected in the change in EDSS. [redacted] The graphical representation below shows the change from baseline in EDSS over time in TEMSO.

[redacted] **Change from baseline in EDSS over time**

[redacted]

Finally, fatigue and measures of health related quality of life (SF-36 and EQ-5D) showed no statistically significant differences between groups in individual trials except minor differences in a single mental health domain; and no meta-analysis was conducted.

Discontinuation rates were similar to placebo in TOWER and TEMSO but more patients discontinued compared to placebo in study
2001. [redacted]. [redacted]

Subgroup results

Subgroup analyses were performed by baseline demographics (gender, race and age), disease characteristics (EDSS strata, relapse history, MS subtype, MRI parameters and prior use of DMT for MS. Based on TEMSO, the manufacturer reported that reductions in ARR and disability progression were consistent across subgroups in favour of teriflunomide with no treatment by subgroup interaction test reaching statistical significance.[13] Analyses in TOWER supplied by the manufacturer confirmed a lack of interaction.

The results for the main outcomes from the main trials by HA and RES subgroups are shown in Table 6.

Table 6: HA and RES subgroup results

	Whole population	[redacted]	[redacted]
ARR	**RR vs. placebo**	[redacted]	[redacted]
TEMSO	RR 0.69 (95% CI: 0.58 to 0.81)	[redacted]	[redacted]
TOWER	0.637 (95% CI: 0.512 to 0.793)	[redacted]	[redacted]
3mth SAD	**HR vs. placebo**	[redacted]	[redacted]
TEMSO	0.70 (95% CI: 0.51 to 0.97)	[redacted]	[redacted]
TOWER	0.685(95% CI 0.467 to 1.004)	[redacted]	[redacted]
Discontinuations	**OR vs. placebo**	[redacted]	[redacted]
TEMSO	0.90 [0.65, 1.25]	[redacted]	[redacted]
TOWER	1.06 [0.77, 1.45]	[redacted]	[redacted]

*HA TEMSO n=21 (2% of total population), TOWER n= 156 (13.3% of total population) but definition of HA less strict in TOWER

**RES TEMSO n=72(96.6% of total population), TOWER =253 (21.6% of total population but definition of RES less strict in TOWER

The results from these post hoc subgroups are of limited
value. [redacted]. The findings are not confirmed in the larger subgroup in TOWER, although those results are confounded by the less specific definition of HA used in that study.

4.2.3 Effectiveness: Active Comparison

TENERE is the only trial that directly compares teriflunomide to another DMT. TENERE is a multinational trial (53 centres in 13 countries) comparing teriflunomide to interferon-beta-1a (Rebif 44µg) in patients with relapsing MS. Patients aged 18 and over (n = 324) were randomised to receive 7mg teriflunomide, 14 mg teriflunomide (111) or 44µg Rebif (104). Teriflunomide was administered once daily and patients were blinded to dosage. Rebif 44µg was administered subcutaneously three times a week on an open label basis. TENERE was designed as a superiority trial to evaluate effectiveness in terms of 'time to failure' and also tolerability. TENERE is as yet unpublished but has been presented at conferences. To be eligible for the trial, patients needed to have RMS according to McDonald (2005) and be ambulatory with an EDSS ≤ 5.5 as for the placebo-controlled trials TEMSO and TOWER. Exclusion criteria were similar, notably patients needed to be relapse-free 30 days prior to randomisation. Prior exposure to Rebif or to another interferon in the previous 3 months was also an exclusion criterion. The study was completed after the last patient reached 48 weeks of treatment so individuals were followed up variously between 48 and 115 weeks.

4.2.3.1 Study design

TENERE is a much smaller trial than the placebo-controlled trials. The study is powered for the composite primary outcome, time to failure; this is in contrast to ARR in TEMSO and TOWER. It may not have adequate statistical power to detect differences in all investigated outcomes. For this reason, claims of 'comparable efficacy' should be treated with some caution. Also the EMA recommends that trials last at least two years to obtain robust data on relapses and sustained accumulation of disability, but not all of the participants were followed up for this time period.[6] TENERE was not a double-blind trial, with neither the care provider nor the patient being blinded to treatment: only the outcome assessment was blinded. Whilst this means the trial can reflect patients' real life response to therapy, it is likely to have biased the evaluation of the primary outcome in favour of teriflunomide.

4.2.3.2 Study population

Table 7: Population of TENERE compared to placebo-controlled trials and RSS

	TEMSO*	[redacted]	[redacted]	Risk Sharing scheme[10]
n	1088	[redacted]	[redacted]	4871
Age mean (SD)	37.9 (8.8)	[redacted]	[redacted]	39.3
% female	72.2	[redacted]	[redacted]	74.9
% white	97	[redacted]	[redacted]	NR
Duration disease since start of symptoms mean (SD)	Median 6.83 yrs.	[redacted]	[redacted]	8.5 (7.4)
Duration	Median 3.5 years	[redacted]	[redacted]	5.2 (5.8)

disease since diagnosis mean (SD)				
EDSS baseline mean (SD)	2.5	[redacted]	[redacted]	3.4 (1.70) RRMS only = 3.1 (1.5)
% EDSS 3.5	NR	[redacted]	[redacted]	NR
Relapses in prior two years, mean (SD)	1 (median in previous 1 yr.))	[redacted]	[redacted]	2.9 (1.2)
% RRMS	91.5%	[redacted]		85.8
Previous treatment with DMT	27.0%	[redacted]	[redacted]	Unclear

The population of TENERE is [redacted] that of the placebo-controlled trials in terms of the population baseline characteristics above.
However [redacted] The proportion of patients with prior use of DMTs [redacted] the placebo-controlled trials.

In terms of the RSS population, TENERE has [redacted] characteristics although [redacted] Importantly, there is some imbalance between the treatment arms in that in TENERE 24% of the Rebif 44µg treatment group had received prior therapy whereas the proportion in the teriflunomide arm was 11.7% .

4.2.3.3 Study outcomes

The primary outcome is time to failure. ARR was not a primary outcome and SAD was not a pre-defined outcome of the trial and instead was defined post-hoc. Relapses and SAD were interpreted in terms of EDSS as for the placebo-controlled trials. In TENERE quality of life was assessed by consideration of satisfaction with treatment medication rather than the usual SF-36.

4.2.3.4 Study internal validity

The table below details the manufacturer's quality assessment of TENERE which has been verified by the ERG. A further critique of the trial is given below.

Table 8: Quality assessment of TENERE trial

[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

4.2.3.5 Effectiveness Results: Active comparator trial

The main results of TENERE are given in the table below. For further detail of results it is necessary to refer to the manufacturer's submission pages 119 to 130, table B6.5.1.

Table 9: Results of TENERE trial

Outcome	***Result (teriflunomide 14mg n=111 vs. Rebif 44µg n=104)***
Primary outcome: Time to failure (confirmed relapse or permanent treatment discontinuation)	42 (37.8%) vs.44 (42.4%) failure HR = 0.861(95% CI: 0.564 to 1.314)
Relapse rate	Adjusted ARR 0.259 (95% CI: 0.153 to 0.438) vs. 0.216 (95% CI: 0.113 to 0.415) Relative risk 1.197(95% CI: 0.623 to 2.299)
Severity of relapse	NR
Proportion relapse-free	[redacted]
Disability: 3 and 6 month SAD	[redacted]
Change in EDSS (mean SD)	[redacted]
Symptoms of MS (FIS)	FIS at 48 weeks (change from baseline) 4.096 vs. 9.099, p = 0.1789
Discontinuation rate	19.8% vs. 28.8%
Treatment Satisfaction Questionnaire for Medication	LS-mean global satisfaction score Week 48: 68.818 (SE 2.782) vs. 60.975 (SE 2.942) p = 0.0162
HrQoL (SF-26 and EQ-5D	Not available

There were no clinically or statistically significant differences between treatments in the primary outcome of time to failure (confirmed relapse or permanent treatment discontinuation). The manufacturer supplied additional tables showing the breakdown of the primary outcome.

Table 10: Analysis of time to failure

	[redacted]	[redacted]
No of patients with primary outcome	42 (37.8%)	44 (42.3%)
Relapse	26 (23.4%)	16 (15.4%)
Permanent treatment discontinuation	15 (13.5%)	25 (24.0%)
Other reason for failure	[redacted]	[redacted]
No. of patients who were censored	[redacted]	[redacted]

[redacted] There were no differences between treatment groups in fatigue associated with MS. Patients in the teriflunomide group had a higher mean global treatment satisfaction than those in the Rebif 44µg group. The data showed that this was mainly due to the side effect profile and convenience of the oral drug. There were no other quality of life measures.

[redacted]. The Figures 3 and 4 supplied by the manufacturer illustrate this.

[redacted]2 [redacted]

[redacted]

[redacted]3[redacted]

[redacted]

4.2.4 Adverse effects of teriflunomide: all trials

The submission did not include a synthesis of adverse effects data that could be readily checked against supporting tables. The summary of adverse effects cited the Summary of Product Characteristics (SPC) for Aubagio®, which could not be checked by the ERG as it is not yet available. A table outlining the adverse events causing discontinuation across the trials would have been informative.

The manufacturer reported that they had constructed two pools of data. Pool A consisted of the placebo controlled parts of Study 2001 and TEMSO. Pool B consisted of Study 2001(both placebo-controlled and extension), TEMSO and its extension, and the teriflunomide arm of TENERE (see 4.2.3). The manufacturer submitted only Pool A as this was comprised of placebo-controlled data. However, it is unclear why safety data from the placebo-controlled TOWER trial was not included. Furthermore, the trials in Pool A lasted six months and two years respectively and therefore there are no long-term pooled safety results included in the submission.

In terms of meta-analysis, the manufacturer pooled TEMSO, TOWER and Study 2001 to obtain data on discontinuation due to adverse events.

Table 11 details rates of adverse events across the trials.

Table 11: Adverse events in placebo trials

Trial	TEMSO	[redacted]	Study 2001	Meta-analysis	Study 2001 + Study 2001 extension	[redacted]
Adverse events	87.5% vs. 90.8%	[redacted]	NS	Not done	100%	[redacted]
Serious adverse events	12.8% vs. 15.9%	[redacted]	NS	Not done	28.8%	[redacted]
Discontinuation due to adverse events	10.9% vs. 8.1%	[redacted]	14.2% vs. 7.0%	[redacted]	19.7%	[redacted]
Mortality	0	[redacted]	0	Not done	1 death (underlying disease and concomitant medications may have been contributory)	[redacted]

Occurrence of adverse events and serious adverse events was similar in teriflunomide and placebo groups in TEMSO and TOWER and results were not pooled in meta-analysis for these outcomes. Overall, meta-analysis of all three placebo trials showed a significantly higher proportion of patients withdrew due to adverse events in the teriflunomide treatment groups[redacted], and there were no deaths in the other placebo-controlled trials.

[redacted]The occurrence of any adverse event remains high in the extension trials. However [redacted]the Study 2001 extension and the TEMSO extension in terms of rates of occurrence of serious adverse events and discontinuation due to adverse events[redacted]

Adverse event data relating to TENERE are presented below as these were not pooled with any other data.[redacted]Table 12: Adverse events in TENERE

	Teriflunomide vs. Rebif 44µg
Adverse effects	92.7% vs. 96.0%
Serious adverse effects	5.5% vs. 6.9%
Discontinuation due to adverse events	10.9% vs. 21.8%
Mortality	No deaths occurred

A similar proportion in each treatment group experienced adverse events and serious adverse events. A greater proportion of those receiving Rebif 44µg discontinued due to adverse events. Again though, a greater proportion of the Rebif 44µg group had previously taken DMTs and the impact of this difference is unknown. No deaths occurred in TENERE.

A breakdown of the adverse event data in TENERE was provided by the manufacturer and is shown in the table below. It can be seen that several adverse events are [redacted]

Table 13: Breakdown of adverse events in TENERE

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

4.2.4.1 Long-term safety of leflunomide

The manufacturer includes in the submission a short section on the long-term safety of leflunomide (see page 190 of the submission)., pointing out the acceptable safety profile of this metabolite of teriflunomide, which has been licensed for the treatment of rheumatoid and psoriatic arthritis for over a decade. As the EMA has granted new active substance status to teriflunomide, the ERG is unclear about the relevance of this experience.

4.3 Additional trials

The ERG did not locate any further relevant trials that the manufacturer failed to identify. The manufacturer indicated that the trials TOPIC and TERACLES were excluded. The ERG agree that

TOPIC is not relevant to the decision problem as it is a population of CIS patients with Clinicall Isolated Syndrome (CIS) only, and that TERACLES is not relevant as it is a trial of teriflunomide administered in combination with a beta interferon.Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

4.3.1 Searches and trial selection

The searches for trials to be included in the MTC were those described in section 4.1.1. As such they were appropriate and well documented strategy.

The main selection criteria for trials for the MTC were:

- Adults with either RRMS or progressive MS (including SPMS or PRMS), of which 80% have RRMS. Trials with mixed MS patient populations with less than 20% PPMS or clinically isolated syndrome (CIS) patients also were included.
- Single- or double-blind randomised controlled trials (RCTs) of at least six months duration and open-label extensions of these trials.
- Patients in at least one arm of the trial received one of the following DMTs (including pooled doses of the same drug). Comparators included placebo or any of these same selected DMTs (including the same drug at a different dose or formulation). In the case of beta interferon, comparators were split by brand and dose. The choice of comparators was to include all possible MS DMTs: alemtuzumab; dimethyl fumarate (BG00012); daclizumab; fingolimod; glatiramer acetate; Interferon-1a; Interferon-1b; laquinimod; mitoxantrone; natalizumab; rituximab; teriflunomide.
- The study-reported relapse, disability progression, magnetic resonance imaging (MRI), or safety outcomes of interest.

Trials whose recruitment was prior to the year 2000 were not included in the base case MTC but were included in an ‘All Years’ sensitivity analysis.

The ERG considers these inclusion criteria appropriate. The inclusion of a wide range of DMTs and all doses thereof studied, whether licenced for use or not, is appropriate in an MTC as it generates the most complete network for each outcome and provides the greatest amount of data for the placebo and other standard comparators. However, not all comparisons are actually of direct relevance to the use of teriflunomide in the NHS at the present time, due to available marketing authorisations and NICE guidance. There is some lack of clarity in that despite the fact that some unlicensed doses of drugs are included in the MTC, 11 studies are excluded for being of ‘Not approved dosing or scheduling’.

The ERG considers that the restriction of the base case network to trials where patient recruitment was from the year 2000 onwards was not appropriate. The reason given for this restriction was that

ARRs have fallen in clinical trials and hence older trials will have higher baseline rates than more recent trials. The cut off at the year 2000 was selected because of the introduction of the McDonald diagnostic criteria in 2000 to replace the Poser criteria. Compared with the Poser criteria, the McDonald criteria generally identify patients much earlier in the disease course, and with less frequent or severe symptoms; therefore, a change in MS diagnostic criteria may have considerable impact on the baseline disease activity of the patient population under study. However, 5 out of 30 trials included in the base case network used the Poser criteria (including the Phase II trial of teriflunomide) (All the 6 excluded trials used Poser criteria). The publication of the CHAMPS study in 2001[14] was also cited as a reason for this restriction. This restriction was supported by Key Opinion Leaders consulted by the manufacturer and by a comparison of baseline relapse rates and treatment effects pre and post 2000. However, the ERG suggested in their clarification requests that the All Years MTC with baseline relapse rate included as a covariate would be more appropriate, as it would include all the data but take any heterogeneity in baseline ARR into account. However, the manufacturer was unable to run this analysis in time for the ERG's report schedule.

One major effect of the restriction on trial inclusion was that all the placebo controlled trials of beta-interferons and of glatiramer acetate were excluded, although one direct placebo comparison of glatiramer acetate was included due to its being a comparator in a placebo controlled trial of dimethyl fumarate. This is potentially important as in the manufacturer's submission beta-interferons and glatiramer acetate are the direct comparators for teriflunomide The impact of this restricted analysis is discussed in the MTC results section.

4.3.2 Missing trials

The flow diagram provided on page 141 of the submission appears to include a discrepancy: 52 trials are include in the qualitative synthesis and of these 11 are excluded. However, the total number of trials remaining for the MTC All years network is 36 not 41. The ERG assumes the unaccounted for 5 trials were those that were excluded because their population was less than 80% RRMS.

Given the large number of treatments included in the manufacturer's MTC (base case and All years) it was not possible to re run and rescreen the searches in order to check that they had not missed relevant trials or had excluded any that should have been included. Instead the ERG compared the identified studies for the older treatments (beta-interferons and glatiramer acetate) with those included in existing systematic reviews[15, 16]and also previous submissions to NICE (Fingolimod).

Excluded from the base case but included in the 'All Years' analysis are three important trials of the beta-interferons and three smaller trials of glatiramer acetate (Table 14).

Table 14 Trials excluded from the base case but included in the 'All Years' analysis

Trial	Comparison	N	Duration (months)
PRISMS[17]	Rebif 44 and 22 vs. placebo	560	24
MSCRG[18]	Avonex 30µg vs. placebo	301	24
IFNB MS study[19]	Betaferon 250 vs. placebo	247	24
Bornstein 1987[20]	glatiramer acetate vs. placebo	50	24
Copolymer 1 MS Study[21]	glatiramer acetate vs. placebo	251	24
European/Canadian glatiramer acetate study[22]	glatiramer acetate vs. placebo	239	9

These were included in the MTC that formed part of the submission to NICE of Fingolimod for RRMS.

In the All Years analysis the submission for teriflunomide has included the same trials as in the MTC in the submission to NICE for the appraisal of fingolimod with the exception of 4 trials, two of which compared the licensed dose of Betaferon with an unlicensed dose, one included only a 90 day treatment period and therefore was too short a study. The one study that appears to be missing from the all years MTC is the INCOMIN study,{Durelli, 2002 #51 which compared Avonex 30µg with Betaferon 250µg (n=188) in a 2 year-long study; its primary outcome was proportion of patients relapse free at 24 months but it also reported ARR and SAD (but 6 month confirmation only). It is unclear what impact on the All Years results for ARR the inclusion of this trial would have had on the teriflunomide comparisons: examination of the network suggests that as there is no direct link between Betaferon 250µg and teriflunomide and therefore this trial would have had only a small impact on the MTC results.

A search by the ERG for recent trials of beta-interferons and glatiramer acetate identified two possible trials for inclusion: Calabrese 2011{Calabrese, 2012 #52} was a 24 months study of Avonex 30µg vs. Rebif 44µg vs. glatiramer acetate, but the only outcome it reported relevant to the submission was withdrawals, and therefore the omission of this trial is not of great importance; and IMPROVE 2010 [23], a comparison of Rebif with placebo that used only a 16 week treatment period and was therefore not directly relevant.

4.3.3 Included trials

The teriflunomide submission MTC included a large number of trials: 30 in the base case and 36 overall. The All Years network for the primary outcome of ARR included 32 trials. Of these 4 used data at 12 months and a further 7 used data for less than 12 months. 12 months is a short duration for the assessment of an infrequent event such as MS relapse. The ERG did not rerun the MTC analysis to check what would be the impact of the removal of the 11 short trials. The network diagrams are provided in the manufacturer's submission (Figures B6.7.2 to B6.7.6, pages 154-158, and Additional Appendices Figures AA2.1 to AA2.3).

The All Years network for the key outcome in the model - 3mth SAD - included 20 trials. A number of the trials not included in this outcome network were of short duration (12 months or less) and therefore their omission (whilst not pre-specified for this reason) is appropriate. Other potentially relevant trials did not report 3mth SAD as an outcome. BECOME[24- a 24 month study of Betaferon 250µg vs. glatiramer acetate 20mg (SAD not an outcome){Mikol, 2008 #38]study of Rebif vs. glatiramer acetate 20mg (only reported post-hoc 6mth SAD); CombiRx[25] – a 36 month trial that included a comparison of Avonex 30µg with glatiramer acetate 20 mg; (only reported 6mth SAD); BRAVO[26] – a 24 month trial of laquinimod vs. Avonex 30µg vs. placebo; (SAD not an outcome); The MSCRG trial [18] – a 24 month comparison of Avonex 30µg with placebo (only reported 6mth SAD). It is worth noting the three of these trials did report 6 month SAD (the more robust outcome):REGARD,[27] CombiRx,[25], and the MSCRG trial.[18]

Of the included trials for 3 month SAD, two used data at 12 months and a further one used data for less than 12 months. 12 months is a short duration for the assessment of an infrequent event such as confirmed progression associated with MS relapse. The ERG did not rerun the MTC analysis to check what would be the impact the removal of the 11 short trials?

In summary, the MTC was inclusive with very few trials omitted. Whilst the Manufacturer's concerns regarding the inclusion of older trials are to some extent justified and neither the base case nor the All Years analysis are optimal, omission of the beta-interferon vs. placebo trials from the base case reduced the reliability of the results.

4.4 Critique of the indirect comparison and/or multiple treatment comparison

4.4.1 Methods

In their submission the manufacturer provided very few details regarding the methods of the MTC. The MTC utilised a Bayesian random effects model that employed a vaguely informative normal prior with a uniform standard deviation, the value of which was 1.0 for all outcomes. The models were run appropriately and the submission presented evidence of their convergence. The submission did not report how 3 or more arm trials were dealt with in the model, nor did it report statistics of model fit. Following the ERG's request for further details of the methods the manufacturer issued the DIC

statistics for the model (for fixed effect and well as random effects) and confirmation that the model included relative, not absolute effects. However, no details of residual deviance, which is an important indication of model fit were presented: the model fit is the extent to which the model explains the data. If a model is a good fit then the residual deviance should be close to the number of data points in the data set.

The DIC statistics provided for the MTC analyses were: for ARR , RE 608.4 and FE 607.6; for 3 mth SAD, RE 7.975 and FE 6.333; and for Discontinuations, RE 474.3 and FE 472.3. The similarities between the DIC for the random and fixed effect models indicates that the random and fixed effect models are equally good fits to the data. Further discussion refers to the results from the random effects model only.

Examination of the code by the ERG found it to be correct and appropriate. No attempt was made to rerun the network analyses.

4.4.1.1 Heterogeneity

Heterogeneity was explored informally as is appropriate in an MTC although further investigation of the impact of potential confounders should have been undertaken. Trial and population characteristics for all the trials in the MTC were presented in the manufacturer's submission (tables B6.7.1 and AA2.1) and are summarised in Tables 15 and 16 below. The ERG would like to have seen some covariate analyses using the all years network to control for potential confounders: baseline ARR and baseline disease duration, baseline EDSS (though this was adjusted for in the teriflunomide analyses of ARR); proportion of patients previously treated with a DMT. These are factors which were accounted for either by adjustment in analysis of ARR or by exclusion of pre-2000 trials from the MTC base case.

Table 15: Range of potential confounding factors across the trials in the MTC (base case)

MS Diagnosis Criteria	Population	Recruitment Period (earliest to latest across all trials)	Age (SD), Years in any one treatment arm - range	% Male in any one treatment arm	Baseline EDSS	Previous Relapse Rate in any one treatment arm - range	Prev. Treated	Disease Duration, Years* in any one treatment arm - range
McDonald (23) Poser (5) NR (2)	100% RRMS – 23 trials (77%) 95 - 99% - 3 trials* 90% - 3 trials$ 80% RRMS – 1 trial	2001–2011	31.9 (SD 8.0) to 42.4 (SD not reported) or 41.5 SD 8.5)	17% to 44.6%	1.8 (SD 1.7) to 2.9 (SD 1.3) or 3.2 (SD not reported)	Over 1 yr. 1.0 to 2.4 Over 2 yrs. 1.9 to 2.6	0% 3 trials >0% to 50% - 6 trials >50% - 99% 5 trials NR 15 trials	2.0 (SD 1.3) to 10.8 (SD 8.2) (in 8 not reported)

*2 trials (both teriflunomide just less than 100% - TOWER 97% and TENERE 99%; $Phase II teriflunomide trial

The ERG identified some clinical heterogeneity across the trials in the base case MTC (Table 15): not all used the McDonald diagnostic criteria despite the year 2000 cut off; the majority of trials (77%) included only RRMS patients, though interestingly none of the teriflunomide trials did; previous DMT use was not reported in half the trials; and only 3 trials (10%) had a previously untreated population.

Table 16: Range of potential confounding factors across the trials in the MTC (All Years analysis))

MS Diagnosis Criteria	Population	Recruitment Period	Age, Years	Male	Baseline EDSS	Previous Relapse Rate	Prev. Treated	Disease Duration, Years*
McDonald (23) Poser (11) NR (2)	100% RRMS – 29 trials (81%) 95 - 99% - 3 trials* 90% - 3 trials$ 80% RRMS – 1 trial	1988–2011	30.0 (SD not reported) or 31.9 (SD 8.0) to 42.4 (SD not reported) or 41.5 SD 8.5)	17% to 44.6%	1.8 (SD 1.7) to 2.9 (SD 1.3) or 3.2 (SD not reported)	Over 1 yr. 1.0 to 2.4 Over 2 yrs. 1.9 to 3.9	0% 5 trials >0% to 50% - 6 trials >50% - 99% 5 trials NR 19 trials	2.0 (SD 1.3) to 10.8 (SD 8.2) (in 8 not reported)

Clinical heterogeneity was similar in the All Years network. As would be expected the proportion of studies using the Poser diagnostic criteria was higher and the recruitment range was wider. The previous relapse rate over 2 years included a higher upper value 3.9) as did the ARR at the end of trial. Age range, % male, baseline EDSS range and disease duration range were not different from the base case.

4.4.1.2 Consistency

Formal statistical consistency analysis was not reported in the submission In order to examine the face validity of the MTC the ERG requested a full set of the MTC results, both for the base-case analysis and for the All years analysis. The ERG were provided with the full base-case results; unfortunately

the full results for the All Years analysis were not provided. Examination of the full set of base-case results enabled the ERG to compare the MTC generated results with the available direct comparisons. Whilst most direct results were derived from single trials some were derived from meta-analyses of trials. The ERG has been unable due to time constraints to verify these individual results nor the appropriateness of any meta-analyses (other than that of teriflunomide trials). Direct comparisons with placebo were included for most treatments but notably there were none included for the beta-interferons: this was due to the exclusion of the key beta-interferon placebo controlled trials from the base case analysis. The comparisons between indirect and available direct evidence (See Appendix Table) showed that the MTC results were not notably different from the trial results, indicating that the MTC model results are consistent. However, small changes in key comparisons have important impact in the economic model; these are discussed further in the MTC results section 4.4.2).

4.4.2 Results of MTC

The results of the MTC base case and All years are presented in Tables 17 to 19 by outcome (ARR, 3mth SAD and Discontinuations.

Table 17: MTC results ARR

	Base case analysis	All Years analysis	Direct results available*
Comparison	**Rate ratio [95% CrI]**	**Rate Ratio [95% CrI]**	
Teriflunomide 14mg vs. Placebo	0.67 [0.57, 0.77]	[redacted]	O'Connor P, 2006: 0.68 [0.36, 1.27] TEMSO: 0.69 [0.58, 0.81] TOWER: 0.64 [0.53, 0.77] Direct MA: 0.664 [0.589, 0.749]
Betaferon 250µg vs. Placebo	0.68 [0.52, 0.88]	[redacted]	
Avonex 30µg vs. Placebo	0.78 [0.67, 0.91]	[redacted]	
Rebif 44 mg vs. Placebo	0.62 [0.51, 0.76]	[redacted]	
GA 20 mg vs. Placebo	0.64 [0.53, 0.76]	[redacted]	CONFIRM: 0.72 [0.61, 0.87]
Fingolimod 0.5 mg vs. Placebo	0.46 [0.40, 0.54]	[redacted]	Kaposi L, 2010: 0.45 [0.37, 0.54] Saida, 2012: 0.51 [0.27, 0.95] Calabresi, 2012: 0.52 [0.43, 0.64] Direct MA: 0.486 [0.425, 0.556]
Natalizumab 300 mg vs. Placebo	0.31 [0.25, 0.39	[redacted]	Polman, 2006: 0.32 [0.27, 0.37]
Teriflunomide 14mg vs. Betaferon 250µg	0.98 [0.73, 1.31]	[redacted]	
Teriflunomide 14mg vs. Avonex 30µg	0.86 [0.69, 1.05]	[redacted]	
Teriflunomide 14mg vs. Rebif 44 µg	1.06 [0.84, 1.35]	[redacted]	[redacted]
Teriflunomide 14mg vs. GA 20 mg	1.05 [0.83, 1.31]	[redacted]	
Teriflunomide 14mg vs. Fingolimod 0.5 mg	1.45 [1.17, 1.80]	[redacted]	
Teriflunomide 14mg vs. Natalizumab 300 mg	2.12 [1.63, 2.75]	[redacted]	

*For complete table of direct vs. indirect comparisons available see Appendix TableXX; GA glatiramer acetate

Adjusted ARR rates from the teriflunomide trials were used in the MTC, but it is unclear whether the ARR values from other trials were adjusted or not. The ERG is unsure what impact this will have had on the results. Comparison of the ARR results direct from trials with those from the MTC found that results for almost all comparisons were very similar.

The MTC results for ARR indicate statistically significant treatment benefit for all DMTs. When compared with the other DMTs, teriflunomide was similar to Betaferon 250µg, slightly better than Avonex 30µg (but not statistically significantly so), slightly worse than both Rebif 44µg and glatiramer acetate (but again not statistically significantly so), and significantly less effective than fingolimod and even more so than natalizumab.

Table 18: MTC results 3mth SAD

	Base case analysis	All Years analysis	Direct results available*
Comparison	**Hazard ratio [95% CrI]**	**Hazard Ratio [95% CrI]**	**Hazard Ratio [95% CrI]**
Teriflunomide 14mg vs. Placebo	0.71 [0.53, 0.92]	[redacted]	TEMSO: 0.7 [0.51, 0.96] TOWER: 0.68 [0.47, 1] Direct MA: 0.694 [0.544, 0.886]
Betaferon 250µg vs. Placebo	1.21 [0.68, 2.16]	[redacted]	IFNB Study, 0.68 (95% CI 0.40, 1.17)
Avonex 30µg vs. Placebo	0.91 [0.61, 1.33]	[redacted]	Data only for 6mth SAD
Rebif 44 µg vs. Placebo	0.79 [0.51, 1.24]	[redacted]	PRISMS, 0.65 (95% CI 0.45, 0.94)
Rebif 22 µg vs. Placebo		[redacted]	
GA 20 mg vs. Placebo	0.93 [0.59, 1.45]	[redacted]	CONFIRM: 0.93 [0.63, 1.37]
Fingolimod 0.5 mg vs. Placebo	0.75 [0.58, 0.96]	[redacted]	
Natalizumab 300 mg vs. Placebo	0.58 [0.4, 0.84]	[redacted]	Polman, 2006: 0.58 [0.43, 0.78]
Teriflunomide 14mg vs. Betaferon 250µg	0.58 [0.30, 1.12]	[redacted]	
Teriflunomide 14mg vs. Avonex 30µg	0.77 [0.50, 1.24]	[redacted]	
Teriflunomide 14mg vs. Rebif 44µg	0.90 [0.54, 1.45]	[redacted]	[redacted]
Teriflunomide 14mg vs. Rebif 22µg		[redacted]	
Teriflunomide 14mg vs. GA 20 mg	0.76 [0.45, 1.30]	[redacted]	
Teriflunomide 14mg vs. Fingolimod 0.5 mg	0.95 [0.64, 1.35]	[redacted]	
Teriflunomide 14mg vs. Natalizumab 300 mg	1.22 [0.77, 1.94]	[redacted]	

*For complete direct vs. indirect comparisons available see Appendix Table; GA glatiramer acetate

[redacted]

[redacted]The two trials which will have had the greatest influence of the teriflunomide vs. Rebif 44µg comparison in the base case MTC are the TRANSFORMS study,[28] a comparison of fingolimod 0.5 mg with Avonex 30µg, and a direct comparison of Rebif 44 and Avonex 30µg in beta-interferon naive patients.[29] The TRANFORMS' population comprised around 55% previous DMT, which is quite high compared to other trials, but not unusual across all the trials: TEMSO (27%) and TOWER (33%) and TENERE 19%. This trial may have underestimated the effect of Avonex 30µg, which then may have impacted on the Avonex 30µg comparison with Rebif 44µg, and may possibly be expected to overestimate the relative effect of Rebif 44µg compared to teriflunomide.. However, comparison of the direct (0.87 [0.58, 1.31]) and indirect (0.87 [0.61, 1.25]) estimates for Rebif 44µg and Avonex 30µg showed no difference.

The MTC results for 3 month SAD are more difficult to interpret than those for ARR, being less consistent and based on a much smaller network due to the failure of many trials to report 3 month SAD as an outcome. Compared with placebo most DMTs have a small, not statistically significant benefit on 3 month SAD, with only teriflunomide, fingolimod and natalizumab achieving statistical significance. When compared with the other DMTs teriflunomide is better than them all except for the base case estimate compared with natalizumab, although no result is statistically significant. Also the effect size for teriflunomide compared with placebo for 3 month SAD reported in the MTC was greater (and more precise) than the placebo comparisons of all the individual beta-interferons and glatiramer acetate. Thus suggesting, on average, that teriflunomide was more effective than all existing beta-interferons and glatiramer acetate – albeit that this was highly uncertain and not statistically significant. Overall, these results seem somewhat at odds with the ARR results; it is unclear how a greater impact on disability progression might be achieved in the absence of any benefit in terms of relapse rate.

4.4.2.1 Discontinuations

Table 19: Table MTC results Discontinuations

	Base case analysis	All Years analysis	Direct results available*
Comparison	**Hazard ratio [95% CrI]**	**Hazard Ratio [95% CrI]**	**Hazard Ratio [95% CrI]**
Teriflunomide 14mg vs. Placebo	[redacted]	[redacted]	O'Connor P, 2006: 3.8 [1.15, 12.58] TEMSO: 0.9 [0.65, 1.25] TOWER: 1.07 [0.79, 1.45] Direct MA: [redacted]
Betaferon 250µg vs. Placebo	[redacted]	[redacted]	
Avonex 30µg vs. Placebo	[redacted]	[redacted]	
Rebif 22 µg vs. Placebo		[redacted]	
Rebif 44 µg vs. Placebo	[redacted]	[redacted]	
GA 20 mg vs. Placebo	[redacted]	[redacted]	CONFIRM: 0.59 [0.43, 0.81]
Fingolimod 0.5 mg vs. Placebo	[redacted]	[redacted]	Kaposi L, 2010: 0.61 [0.44, 0.85] Saida, 2012: 1.59 [0.53, 4.82] Direct MA: 0.848 [0.348, 2.068]
Natalizumab 300 mg vs. Placebo	[redacted]	[redacted]	Polman, 2006: 0.81 [0.54, 1.2]
Teriflunomide 14mg vs. Betaferon 250µg	[redacted]	[redacted]	
Teriflunomide 14mg vs. Avonex 30µg	[redacted]	[redacted]	
Teriflunomide 14mg vs. Rebif 22µg		[redacted]	
Teriflunomide 14mg vs. Rebif 44µg	[redacted]	[redacted]	[redacted]
Teriflunomide 14mg vs. GA 20 mg	[redacted]	[redacted]	
Teriflunomide 14mg vs. Fingolimod 0.5 mg	[redacted]	[redacted]	
Teriflunomide 14mg vs. Natalizumab 300 mg	[redacted]	[redacted]	

*For complete direct vs. indirect comparisons available see Appendix Comparison of the discontinuations results direct from trials with those from the MTC found that results for almost all comparison were very similar, though for three comparisons of dimethyl fumarate and two of fingolimod the estimate of effect had change by around 0.3 units (see Appendix table); GA glatiramer acetate

4.4.2.2 [redacted]

[redacted]

Rebif 22 µg dose

The base case MTC did not allow the calculation of relative effects for Rebif 22µg because no trials that included this dose were in the base case network. Trials that reported 3mth SAD and discontinuations were included in the All Years networks and for these two outcomes the manufacturer used the ratio of the base case estimates with the All Years estimates for Rebif 44µg to generate estimates for Rebif 22 µg (Tables B6.7.25 in the submission). The inferred rates reported in Table B6.7.25 do not seem consistent: that given for 3 mth SAD is for Rebif 22µg vs. placebo whilst that for discontinuations is for teriflunomide vs. Rebif 22µg. The ERG calculated teriflunomide vs. Rebif 22µg and Rebif 22µg vs. placebo for 3mth SAD and for discontinuations are given in Table XXX below.

Table 20: ERG calculated 3 month SAD and discontinuations

	Base case analysis		**All Years analysis**	
	Teriflunomide 14mg vs.	**Vs. Placebo**	**Teriflunomide 14mg vs.**	**Vs. Placebo**
3mth SAD (Hazard ratio [95% CrI])				
Rebif 22 µg	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44 µg	0.90 [0.54, 1.45]	0.79 [0.51, 1.24]	[redacted]	[redacted]
Discontinuations (Odds ratio [95% CrI])				
Rebif 22 µg	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44 µg	[redacted]	[redacted]	[redacted]	[redacted]

If the vs. placebo results had been calculated from direct comparison of Rebif 44 and 22 (PRISMS trial)[30] the results would have been: HR for SAD Rebif 22 vs. placebo [redacted] ; OR for discontinuations Rebif 22 vs. placebo = [redacted].

4.4.3 Main conclusions of critique of MTC

The MTC analysis was appropriate in that it included all relevant comparators of teriflunomide, though not all comparisons are actually of direct relevance to the use of teriflunomide in the NHS at the present time. Only one potentially relevant trial was missed form the MTC analysis, which could have informed the ARR results but, it would probably not have had a major impact on the estimates of the relative effectiveness of teriflunomide. The ERG has one major criticism which is of the decision to exclude pre-2000 trials from the base case analysis. In the opinion of the ERG some covariate analyses using the All Years network to control for potential confounders would have been more appropriate and would have avoided the loss of the placebo comparisons for the beta-interferons. The methods of the MTC were appropriate, except that neither formal assessment of consistency nor values of residual deviance were reported. Informal checks for consistency by the ERG did not identify major problems with consistency except with for the impact on the comparison of Betaferon 250µg with placebo and to a lesser extent the comparison of Teriflunomide with Rebif 44µg, generated by the smaller base case network for 3 month SAD. This inconsistency may have contributed to the favourable results for teriflunomide generated by the MTC (particularly the base case analysis). Because the relative effect of teriflunomide in terms of 3 month SAD is such a key driver in the model the MTC result for these are very important. The inconsistency was less marked with the All Years network and therefore the ERG considers this more complete network is to be preferred.

4.5 Additional work on clinical effectiveness undertaken by the ERG

The ERG did not undertake any further additional work relevant to the clinical effectiveness.

4.6 Conclusions of the clinical effectiveness section

The manufacturer presented a well-conducted systematic review of RCTs involving any treatment for RRMS / PMS in adults. Three placebo-controlled trials were identified for the review (TEMSO, TOWER and Study 2001) which were pooled in meta-analyses to produce summary results. The two key placebo studies are TEMSO and TOWER, both of which are well-conducted, international trials with over 1000 participants. These were supplemented by two extension studies to TEMSO and Study 2001 (without a placebo-controlled arm) which provided longer term safety and efficacy data. The submission also included a trial that directly compared teriflunomide to another DMT (Rebif 44µg). To supplement the direct evidence, the manufacturer undertook a MTC to assess the effectiveness of teriflunomide compared to other DMTs. The results of the MTC informed the economic model assessing the cost-effectiveness of teriflunomide.

Although the placebo-controlled trials were well-conducted they were of short duration given the lifelong duration of MS and the infrequency of relapses. The active comparator trial was smaller and

underpowered for the main outcomes of interest to this submission (ARR and SAD) and it was not double-blinded and therefore the trial could have been biased in favour of teriflunomide.

Teriflunomide significantly reduced adjusted ARR and 3 month SAD compared to placebo. Teriflunomide did not significantly reduce 6 month SAD in the two main trials, TEMSO and TOWER. The placebo controlled trials failed to demonstrate significant benefits of teriflunomide in terms of [redacted], Fatigue Impact Scale (FIS), and Health-related quality of life(HrQoL) (SF-36 or EQ-5D). Although severity of relapse and freedom of disease activity showed superiority of teriflunomide over placebo but these were not pre-specified endpoints of the trials, used surrogate outcomes and no meta-analysis was undertaken.

There were no significant differences between teriflunomide and Rebif 44µg for the primary outcome of time to failure (confirmed relapse or permanent treatment discontinuation). There were no significant differences in ARR [redacted] between teriflunomide and Rebif 44µg in TENERE.

The effect size teriflunomide compared with placebo for 3 month SAD reported in the MTC was greater (and more precise) than the comparisons between all the individual beta-interferons/glatiramer acetate and placebo, raising questions about the reliability of the MTC results.

The submission did not include a synthesis of adverse effects data that could be readily checked against supporting tables. The relatively short duration of the placebo-controlled trials impacts on assessment of any differences in mortality and rarer adverse events. Meta-analysis of the three placebo-controlled trials showed that patients [redacted] to discontinue due to adverse events compared to placebo. In TENERE, although a greater number of participants were found to discontinue due to adverse events, this needs to be interpreted in the light of a higher number of patients in the Rebif 44µg arm who had previously taken a DMT. The impact of this difference is unknown.

[redacted]In the TENERE adverse events more common with teriflunomide than Rebif 44µg were diarrhoea and alopecia. Among those more common with Rebif 44µg were influenza-like illness and increase in alanine aminotransferase.

The clinical evidence suggests that teriflunomide appears to have some benefits compared to placebo, notably for short-term relapse rate but its role in preventing accumulation of disability is less clear.

There is a lack of reliable direct evidence comparing teriflunomide with other DMTs; it is not certain if teriflunomide is equivalently effective to Rebif 44µg.

5 Cost Effectiveness

This section focuses on the economic evidence submitted by the manufacturer and the additional information provided following the ERG points for clarification. The submission was subject to a critical review on the basis of the manufacturer's report and by direct examination of the electronic version of the economic model. The critical appraisal was conducted with the aid of a checklist to assess the quality of economic evaluations and a narrative review to highlight key assumptions and possible limitations. Section 6 presents additional work undertaken by the ERG to address some remaining uncertainties.

The manufacturer's initial economic submission included:

1. A description of the systematic search strategy used to identify existing cost-effectiveness studies for teriflunomide in the treatment of relapse-remitting multiple sclerosis (RRMS) (Manufacturer's Submission, Section 7.1) with details in a separate appendix (Manufacturer's Submission, Appendix 10).
2. A report on the *de novo* economic evaluation conducted by the manufacturer. The report described the technology; comparators and patient population (Manufacturer's Submission, Section 7.2); clinical parameters and variables (Manufacturer's Submission, Section 7.3); the assumptions and sources of evidence used to assess quality of life (Manufacturer's Submission, Section 7.4); the resource use and unit cost assumptions and sources (Manufacturer's Submission, Section 7.5); sensitivity analysis (Manufacturer's Submission 7.6); and the base-case cost-effectiveness results (Manufacturer's Submission 7.7).
3. An Excel-based model comprising the manufacturer's electronic economic model. The ERG has noted that the Excel-based model allows the user to produce the results of scenarios that are not presented or discussed within the main submission. The ERG has chosen to report some additional analyses based on clinical relevance or to inform the understanding of the way the model works. Referred to in the report as the manufacturer's original or uncorrected model.
4. A Patient Access Scheme (PAS) submission describing a price reduction agreed by the Department of Health. This submission includes a revised cost-effectiveness analysis reporting an ICER using the PAS price in the manufacturer's base case analysis and some sensitivity and scenario analyses.

In response to the request for clarification made by the ERG, the manufacturer further submitted:

5. A descriptive reply to the ERG's points for clarifications. This is referred to in the following sections as the manufacturer's response.
6. An updated Excel-based model correcting for an error found in the model. This model is referred to in the report as the manufacturer's corrected model.

5.1 ERG comment on manufacturer's review of cost-effectiveness evidence

5.1.1 Searches

The manufacturer's submission described the search strategies used to identify cost-effectiveness studies relevant to this appraisal of Teriflunomide for the treatment of relapsing-remitting multiple sclerosis in adults. Search strategies were only briefly described in the main submission, however full details were provided in the Appendices.

The electronic databases MEDLINE, MEDLINE In-Process (via Embase) and EMBASE(via Embase), EconLit (Ovid), the Cochrane Library (Wiley) including the NHS Economic Evaluation Database (NHS EED). In addition to this, a grey literature search was performed.

Database searches were performed on 10th and 11th October 2012. Search strategies for each database were documented in Table C10.10.2. The searches covered the period 1966 - October 2012 for Medline and Embase, 1961-October 2012 for Econlit and 1968- October 2012 for NHS EED. No language or date limits were applied to the search. The search excluded animal studies as well as the publication types 'letter' and 'editorial', and 'note'.

The searches were appropriate and comprehensive, and included the use of both subject indexing terms and free text searching. Field searching, Boolean operators and truncation were used where required. All NICE required databases were searched, as well as medical society and regulatory body websites.

Methodological search filter adapted from SIGN was included to identify economic studies and utilities in MEDLINE and EMBASE. The adapted economic study design search filter may have excluded potentially useful records from Medline and Embase.

The addition of the following EMTREE terms would have improved the filter used in EMBASE and Medline: 'Economic evaluation' and 'Cost utility'.

An important issue is that the manufacturer didn't adapt the search for Medline, instead they searched Embase and Medline together from the Embase interface. They used the same search and the same methodological filter developed for Embase. The Cochrane Manual advises to adapt the searches to every database.

5.1.2 Conclusions of the cost effectiveness review

The systematic review of the literature resulted in no cost-effectiveness analyses of teriflunomide. Two cost-effectiveness analyses of other treatments were identified to help inform the development of the *de novo* model.

5.2 ERG's summary and critique of manufacturer's submitted economic evaluation

An overall summary of the manufacturer's approach and signposts to the relevant sections in the manufacturer's submission are reported in Table 21 below:

Table 21: Summary of the Manufacturer's economic evaluation (and signposts to manufacturer's submission)

	Approach	**Source / Justification**	**Signpost (location in manufacturer submission)**
Model	Markov cohort model that tracks disability progression (using the EDSS scale), the occurrence and severity of relapses and the conversion from RRMS to SPMS.	The structure of the model is similar to previous NICE technology appraisals in MS (TA32 [NICE, 2002], TA127 [NICE, 2007] and TA 254 [NICE, 2011]).	*Sections 7.2.2 and 7.2.3*
States and events	The model includes a total of 20 health states: 10 EDSS states for RRMS and SPMS, as well as death. Progression through these states is irreversible in the base case. The occurrence and severity of relapses, adverse events and discontinuation from treatment was also allowed.	This approach to defining states and events in multiple sclerosis is consistent with previous appraisals.	*Sections 7.2.4 and 7.2.5*
Comparators	beta-interferons (Avonex 30µg, Betaferon 250µg, Rebif 22µg and Rebif 44µg) glatiramer acetate natalizumab fingolimod Blended comparator – weighted average based on current market share of beta-interferons and glatiramer acetate	Based on treatments outlined in the NICE Final Scope.	*Section 7.2.7*
Natural History	Based on the Markov model (discussed above), natural history comprises disability progression, relapse, conversion from RRMS to SPMS and mortality. Natural history was assumed in the submission to represent the course of disease under BSC.	Data for disability progression and conversion were derived from a large observational data set – the London Ontario dataset. Data for relapse rates were calculated using published data (Held, 2005). Data for mortality were derived from national mortality statistics adjusted for the additional risk of mortality for different EDSS states (Pokorski,1997). Natural history data was mainly derived from external observational data sets in the	*Section 7.3.1*

		base case, but data was also available from the clinical trials (TEMSO and TOWER).	
Treatment effectiveness	The model assumes treatments impact on the rate of disease progression and on the occurrence and severity of relapse. Mortality rates are assumed to be treatment independent but because mortality is linked to disability progression there will be an indirect effect of treatments on mortality.	Treatment effectiveness is mainly derived from a mixed treatment comparison (i.e. relapse, progression and withdrawal). The treatment effect on the severity of relapse is estimated from TEMSO and other published sources.	*Section 7.3.1*
Adverse events	Grade III/IV adverse events are included in the model if the probability of occurrence is ≥4%.	Adverse event rates are derived from the placebo controlled studies used in the MTC.	*Section 7.3.1*
Health related quality of life	Separate utility values were assigned to each of the EDSS states as well as the effect of relapses by severity. Utility of caregivers was also considered as a variable dependent on the patient's EDSS state.	The utility values were derived from the published literature, and the disutility data from previous NICE multiple sclerosis submissions.	*Section 7.4.9*
Resource utilisation and costs	The following cost categories were considered in the manufacturer analyses: drug acquisition costs, administration costs, monitoring costs, direct medical and non-medical costs and adverse event costs.	The data sources used included UK reference costs and published literature. The unit cost of drugs was based on the risk sharing scheme where available otherwise the NHS list prices (BNF).	*Section 7.5*
Discount rates	A 3.5% discount rate was employed for both costs and health benefits.	In accordance with the NICE reference case approach.	*Table B7.2.11*
Population and Subgroups	The main population considered all patients with RRMS. Subgroup analysis is undertaken for HA and RES populations.	The main analysis matches the population of the licensed indication. Post-hoc subgroup analyses were performed in response to the scope of the decision problem proposed by NICE.	*Sections 7.2.1 and 7.9*
Sensitivity analysis	Scenario analysis and probabilistic sensitivity analysis (PSA) were undertaken.	Structural and deterministic sensitivity analyses are presented. A scatter plot and a cost-effectiveness acceptability plane are presented for the base case model comparing teriflunomide to the blended comparator.	*Section 7.7.7*

5.2.1 The manufacturer's economic evaluation compared with the NICE reference case checklist

Table 22 summarises the economic submission and the ERG's assessment of whether the *de novo* evaluation meets NICE's reference case and other methodological recommendations.

Table 22: NICE reference case checklist

Elements of the economic evaluation	***Reference Case***	***Included in submission***	***Comment on whether de-novo evaluation meets requirements of NICE reference case***
Comparator(s)	Therapies routinely used in the NHS, including technologies regarded as current best practice	Yes	The cost-effectiveness analysis includes all treatments included in NICE's Final Scope. The main comparator was a market share weighted average of these treatments, excluding BSC, fingolimod and natalizumab.
Type of economic evaluation	Cost-effectiveness analysis	Yes	
Perspective on costs	NHS and PSS	Yes	NHS and PSS costs have been taken into account
Perspective on outcomes	All health effects on individuals	Yes	QALY benefits associated with disability progression of individuals and their caregivers were considered.
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	Time horizon analysed was 50 years
Synthesis of evidence on outcomes	Systematic review and mixed treatment comparison of relative effects.	Yes	A systematic review on treatment effectiveness measures and a mixed treatment comparisons (MTC) were conducted.
Measure of health effects	QALYs	Yes	
Source of data for measurement of HRQL	Reported directly by patients and/or caregivers	Yes	Evidence from the published literature was used to assign a HRQL value to each EDSS state based on the EQ-5D. Disutility of relapses and adverse events was also derived from the published literature. Direct evidence from the TEMSO trial was used in a sensitivity analysis.
Source of preference data for valuation of changes in HRQL	Representative sample of the public	Yes	Utility values were based on EQ-5D estimates reflecting public preferences.

Discount rate	Annual rate of 3.5% on both costs and health effects	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was conducted as well as deterministic and structural sensitivity analyses. Mean increment results for the probabilistic sensitivity analysis were presented as well as graphical results using scatter plots, cost-effectiveness acceptability curves and cost-effectiveness acceptability frontiers.

Abbreviations: HRQL, health related QoL; NHS, National Health Service; PSS, personal social services; QALYs, quality-adjusted life years

5.2.2 Population

The manufacturer's main submission considered adults with RRMS for which teriflunomide was granted market authorisation in March 2013. The patient population treated with teriflunomide 14mg in the phase III pivotal trials (TEMSO, TOWER and TENERE) consisted of 95.8% RRMS patients.

As discussed previously in Section 4 of this report, the RRMS population includes highly active (HA) and rapidly-evolving-severe (RES) patients as well as patients that are neither HA nor RES. In previous NICE appraisals (TA 254) the ERG considered these patients to be sufficiently different that they requested analyses for each of these sub-groups separately. The manufacturer provides analyses in the HA and RES subgroups as suggested (i.e. when comparing against natilizumab in RES patients and fingolimod in patients with HA), but do not provide an analysis in the RRMS non-HA/RES population. This is a potentially important issue since the main focus of the manufacturer's submission was in patients for who beta-interferons and glatiramer acetate were the appropriate comparators (i.e. the non-HA/RES population).

As discussed in Section 4 TEMSO and TENERE trials reported a relatively small proportion of HA or RES patients while TOWER included 156 HA patients and 253 RES patients out of 761 patients on placebo or 14mg teriflunomide.

[redacted] (Manufacturer Submission, Table B6.7.11 and B6.7.12, Table B6.7.17 and B6.7.18) (Points for Clarification, Table

A18-A25). Subgroup analyses excluding HA and RES patients was not explored, but it logically follows from the HA and RES subgroup results that teriflunomide may be less effective (on average) in the non-HA/RES population than in the full RRMS population. It should be noted that the manufacturer's definition of HA and RES differed from what has been previously used by NICE as it did not include any MRI outcomes. Specifically, gadolinium-enhancing lesions were not measured at baseline.

Two additional populations were specified in the NICE Final Scope; SPMS with relapses and PRMS. Since teriflunomide did not receive market authorisation for these populations evidence was not presented by the manufacturer.

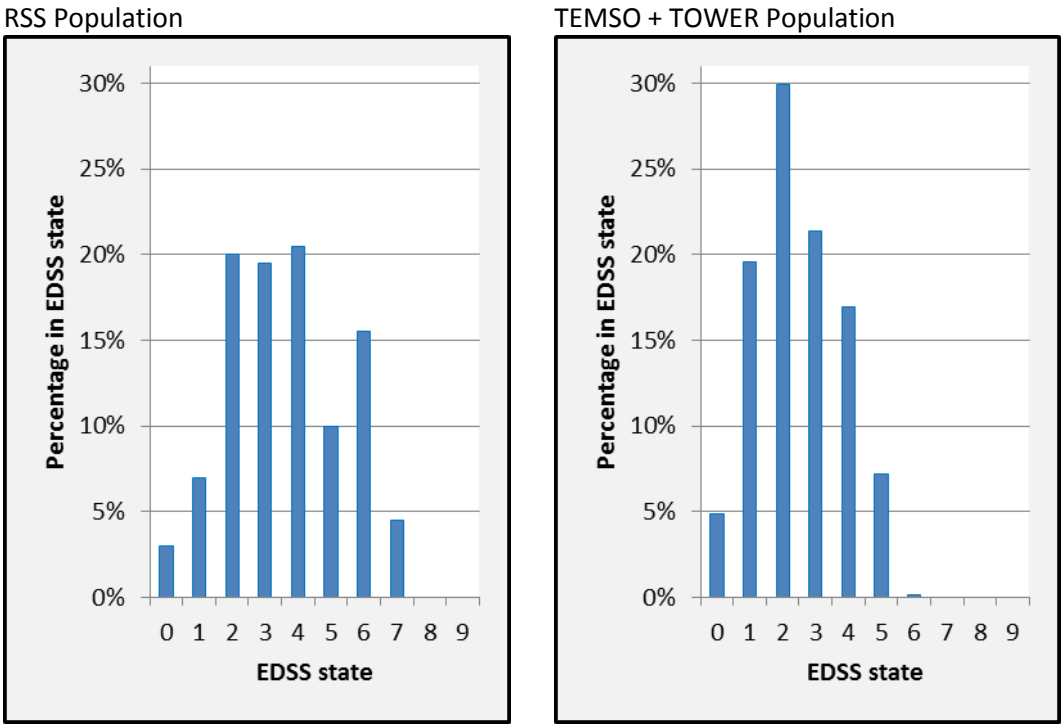
The UK Risk Sharing Scheme (RSS) was used to estimate age, gender and the initial EDSS distribution to define the baseline population characteristics applied in the economic analysis. The RSS population included 85.8% RRMS patients, however the characteristics of this population were applied to analyses representing 100% RRMS patients. The population characteristics of the RSS population, trial populations and modelled population are reported in Table 23.

Table 23: Patient characteristics across studies

Characteristic	RSS Population	TEMSO & TOWER combined population	Base case model
Age, years	39.3	37.9	39.3
Female-to-male ratio	2.98:1	2.52:1	2.98:1
% EDSS ≤3	49.5	75.7	49.5
% RRMS	85.8	95.8*	100

* TEMSO, TOWER and TENERE combined

The patients' initial distribution by EDSS states is more severe in the RSS population than in the trial populations (**Error! Reference source not found.**).



Error! Reference source not found.

Figure 4 Error! Reference source not found.**: EDSS state distributions across studies**

As teriflunomide is expected to be used as first-line therapy, the ERG considered the less severe distribution of the EDSS states from the trial population to be more appropriate for the economic analysis. The trial population also includes fewer patients not considered RRMS. Indeed, the manufacturer's submission itself states that "a lower EDSS at treatment initiation is more representative of the UK target population for teriflunomide which is early in the disease as first line treatment" (Manufacturer's submission, page 237).

5.2.3 Interventions and comparators

The intervention considered in the manufacturer's submission is the licensed dose of teriflunomide, 14mg taken once daily. The TEMSO, TOWER and TENERE studies also evaluated a lower dose of 7mg, this unlicensed dose was not included in the cost-effectiveness analysis.

The comparators in the model included best standard care (BSC), rebif 22µg, rebif 44µg, interferon beta-1a, interferon beta-1b, glatiramer acetate, fingolimod and natalizumab. The manufacturer's base case comparator is a 'blended comparator', excluding BSC, fingolimod and natalizumab. The pooling was based on market share. The blended comparator is calculated as the weighted average of the clinical efficacy, costs and disutility inputs. The manufacturer's justification for using a blended comparator is that there is variation in current practice and previous NICE appraisals have done so (TA 254). However, the manufacturer also reports a fully incremental base case analysis.

Fingolimod and natalizumab were not included in the pooled comparator because their use has been restricted by NICE to HA and RES populations, respectively. Separate comparisons were undertaken with fingolimod and natalizumab in the full RRMS population and using HA and RES subgroups. However, it should be noted that the manufacturer reported that they did not consider either fingolimod or natalizumab to be relevant comparators for teriflunomide and that these subgroup analysis were primarily presented to be in line with the NICE scope. The manufacturer anticipated that teriflunomide would be used in patients with RRMS who would otherwise receive beta-interferons or glatiramer acetate (i.e. the non-HA/RES RRMS population).

The ERG considers the manufacturer's blended comparator to be inappropriate. The issue of blended comparators has been previously discussed by the NICE Decision Support Unit (DSU).[31] The DSU state that the combined comparator (or blended comparator) does not meet the guidelines as the reference case requires 'best practice' to be used as the comparator for appraisal of health technologies. The DSU go on to state "if best practice is defined as the current cost-effective treatment option, the standard approach to assessing the cost-effectiveness would be to consider all treatment options in a single incremental analysis". The NICE methods guide also states that fully incremental analysis should be undertaken.[32]

Further the manufacturer's method for calculating the blended comparator is inappropriate. The manufacturer calculates the blended comparator as the weighted average of each treatment's model inputs i.e. hazard ratio of disease progression, relative risk of annualised relapse rate, etc. The outcomes of the average treatment effects are not the same as the average outcomes of the treatments. This is because of the correlation between costs and benefits in the model. Instead the model outputs (i.e. costs and QALYs) should be averaged to incorporate the potential correlation between the model inputs. This is undertaken by the ERG in Section 6.2.2 to demonstrate the potential effect of this corrected calculation. However, this should not be seen as tacit agreement by the ERG to the use of blended comparators. Finally the manufacturer's calculation of the blended comparator includes a treatment that has worse disease progression than placebo (interferon beta-1b) and treatments that are dominated in the full incremental analysis (interferon beta-1a, interferon beta-1b and Rebif 44µg). The use of the blended comparator masks the effects of changes in the model as different treatments in the blended comparator react differently and sometimes in opposite direction to changes in the model. Its use also hides important model results that can help validate the model and inform clinical practice.

5.2.4 Perspective, time horizon and discounting

The manufacturer's declared perspective is the National Health Service and Personal Social Services. Although this is the stated perspective non-health related costs are also included in the model. These costs and whether or not they meet the NICE reference case are considered further in Section 5.2.9.1

and Section 6.2.5. The modelled time horizon is 50 years with 1 year cycle lengths. In the model 99.5% of the patients are dead at 50 years. Costs and QALYs are discounted at 3.5%. In this model discounting at a higher rate increases the total health effects of treatments. This is contrary to the usual effect of discounting that generally results in lower total costs and health effects with a higher discount rate. This is because a higher discount rates means that future costs and benefits are less important or valuable. Normally future health effects and costs are both positive, so when the discount rate is increased future costs and effects become smaller (i.e. less valuable) and the total costs and health effects also become smaller. In this model the more severe health states have negative health-related quality of life (Section 5.2.8.1) resulting in negative future health effects. By increasing the discount rate the negative future health effects become less valuable (and mathematically larger) which increases the total health effects.

5.2.5 Model structure

In the absence of any previously published cost-effectiveness analysis of teriflunomide the manufacturer undertook a *de novo* economic evaluation based on a decision model. The analysis presented by the manufacturer uses a twenty state Markov model (Figure 6). These states are defined by disability level (EDSS scores 0-9), whether a patient is RRMS or SPMS and death. Patients enter the model with RRMS and can have an EDSS score from 0-7. As previously noted, the base case model uses a baseline EDSS distribution from the UK Risk Sharing Scheme. In each annual cycle patients can remain in the same disability state, progress to a more severe disability state or progress to SPMS. When a patient progresses to SPMS the disability state (i.e. EDSS) also progresses by one level. Patients that have transitioned to SPMS remain in the same disability state or progress in subsequent model cycles. Patients can relapse while in any EDSS state and the probability of relapse is dependent on the severity of EDSS. Each relapse results in additional costs and HRQoL decrement. Patients can die in any state in the model. The probability of death is dependent on the disability state, age and the ratio of females to males. The models cycle length is one year. In the manufacturer's base case patients cannot transition to a lower disability state or regress from SPMS to RRMS.

At any state in the model patients may withdraw from treatment. The model allows up to 3 lines of disease modifying treatment to be administered after which best standard care is administered. After transitioning to SPMS or to an EDSS score greater than 6, patients are no longer treated with disease modifying treatment but receive best standard care (Table 24).

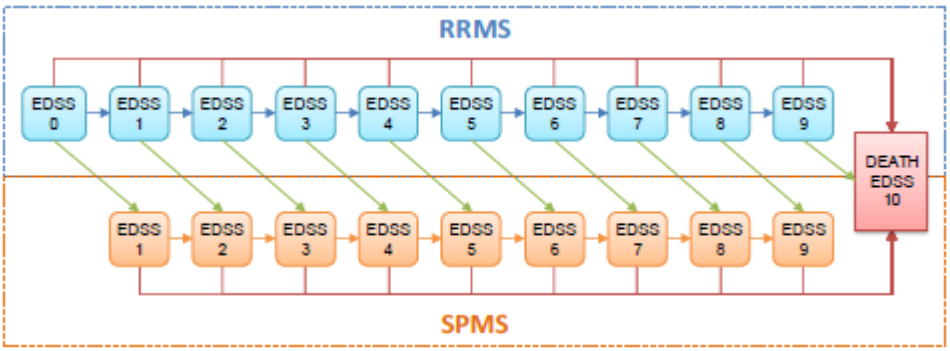


Figure 5: Model structure (Figure B7.2.1, of the manufacturer's submission)*

*The diagram represents patients only transitioning one EDSS state, but the base case model allows patients to transition to any more severe EDSS state.

The model is structurally similar to the models used in previous NICE submissions in MS (TA 32, TA127 and TA 254). The model focuses on six dimensions of the disease:

- Disability progression
- Conversion from RRMS to SPMS
- Relapse
- Mortality
- Treatment discontinuation
- Adverse Events

Each of these will be described in more detail in the subsequent sections on natural history and treatment effects.

Table 24: Structural assumptions of the model

Modelled parameter	Brief description and key assumptions made
Patient entry	The model allows patients to enter the model at RRMS EDSS states 0 to 7 inclusive.
Disability progression	Patient are able to progress within their MS type (ie RRMS or SPMS), to any EDSS state "worse" than their current state. Patients are not able to regress EDSS states. Disability progression occurs independently of the occurrence of relapses, although the reverse is not true. Patients' future disease progression risk is independent of their previous disability progression history.
Maintaining health state	In each period patients may remain in their current EDSS state as well as their current MS type.
Relapse	In any period individuals can experience a relapse, even if maintaining their EDSS score. The frequency of the occurrence of relapses depends only on EDSS state and whether the patient is RRMS/SPMS. A patient who relapses faces the same risk of disability progression as a patient who has not.
SPMS Conversion	RRMS patients face a probability of converting to SPMS in each period; the reverse is not allowed in the model. If conversion occurs the individual also experiences disability progression represented by a 1 point increase in EDSS state. The exception is for RRMS patients with EDSS of 9 that convert to SPMS EDSS 9.
Discontinuation of treatment	Discontinuation of treatment according to the trial withdrawal rates. Conversion to SPMS leads to discontinuation. Disability progression to an EDSS state greater than 6 leads to discontinuation.
Death	In each period an individual faces a risk of death (adjusted for MS).
Extrapolation	A constant treatment effect is applied for as long as patients remain on treatment throughout the 50 year time horizon of the model.

5.2.6 Natural History

The model predicts rates of progression through EDSS states, rates of conversion from RRMS to SPMS, relapse rates and mortality rates. A variety of data sources are used to predict these values for a population not receiving DMT. These values, in combination with the characteristics of the patient population used in the model, are used to define the natural history of the disease. In the submission, natural history is assumed to represent the course of disease under BSC. The treatment effectiveness of teriflunomide and its comparators are defined in comparison to the natural history (BSC) as relative

effects modifying the natural history rates of progression and relapse. These relative effects will be described in more detail in the following section on treatment effectiveness. In this section we describe and critique how the manufacturer has determined the various elements of the natural history as used in the model.

5.2.6.1 Disability Progression

The model considers disability progression in a similar way to previous NICE submissions (TA32, TA127 and TA254), by assuming that both RRMS and SPMS patients experience an underlying risk of disability progression. Disability progression is the main driver of the model and affects relapse rate, conversion to SPMS and mortality.

Disability progression is modelled as progression in EDSS. The EDSS is the standard method for quantifying the severity of MS, and represents a 20 level scoring system (between 0 and 9.5) and is derived by grouping disability in to eight functional systems: pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral and other. The lower EDSS scores of 1.0 to 4.5 represent people who are fully ambulatory, while EDSS scores of 5.0 to 9.5 represent those who have an ambulatory impairment. An EDSS score of 10.0 refers to patients who have died due to MS related symptoms.

The model reduces the 20 EDSS levels to 10, by rounding the half levels. Previous ERG reports comment that the rounding of EDSS states in this way may impact on the assessment of the cumulative probability of sustained progression of disability resulting in an overestimation of the rate of disability progression (TA 127 and TA 254).

Disability progression was derived from a longitudinal dataset of patients with MS from London, Ontario, Canada (Weinshenker et al. 1989). Very little detail about the London Ontario data or the methods used to estimate transitions from the London Ontario data was provided by the manufacturer in the original submission. In response to the ERG's points for clarification the manufacturer provided additional information on the population included in the data set (Manufacturer's response, pg 37). The London Ontario data set contained [redacted] RRMS patients of which [redacted] were 'Active RRMS' and [redacted] were 'Highly Active RRMS'. The mean age was [redacted] with [redacted] females and the mean follow-up was [redacted] years. All patients with only one observation were excluded due to having no transition, leaving [redacted] patients from which transition matrices were calculated (Table 25). The EDSS states were modelled as a homogenous continuous-time Markov process. Transition probabilities were calculated using the msm package from R statistical programming software.

Table 25: Number of transitions per group (Manufacturer's response, Table B.7)

Group	Number of transitions per group								Mean number of transitions
	1	2	3	4	5	6	7	8	
RRMS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pooled Active RRMS*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SPMS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*Including both active and highly active RRMS patients

No data was available from the London Ontario data to inform the transition from EDSS 0 so the model uses data from the placebo arm of the TEMSO and TOWER trials to estimate transitions from EDSS 0. The probabilities estimated from the TEMSO and TOWER trials for transitions from EDSS 0 were combined with probabilities from EDSS 1 to 9 from the London Ontario RRMS dataset to produce the natural history RRMS transition matrix. The London Ontario dataset was also used to estimate transition probabilities between EDSS states 1 to 9 for SPMS patients. No other data was available to estimate transitions for SPMS patients and little has been said about the validity of these transitions however the ERG explores the sensitivity of the model to these transitions in Section 6.2.3.1.

Previous technology appraisals have used the London Ontario dataset for predicting natural history progression without DMT including the original ScHARR model (Chilcott et al. 2003). However, the ERG could not validate the transition matrices or compare to previous submissions since this data is routinely presented as CIC/AIC. Recent technology appraisals have questioned the applicability of this data set given changes in BSC and that it does not collect data on patients regressing. The manufacturer presented information from the TEMSO trial showing a substantial proportion of patients that experienced sustained accumulation of disability (SAD) later regressed (Table 26). This data demonstrates the importance of including regression in the natural history of RRMS patients.

Table 26: Patients experiencing a regression after their first SAD (Manufacturer's response, Table B1)

Status of first Disability Progression	[redacted]	[redacted]	[redacted]	[redacted]
Regression	[redacted]	[redacted]	[redacted]	[redacted]
No regression	[redacted]	[redacted]	[redacted]	[redacted]

The manufacturer undertook a systematic review to identify the most appropriate natural history transition matrix for disability progression for patients receiving no DMT (Manufacturer submission, Additional Appendix 7). The manufacturer claims that the London Ontario data set was found to be the "most appropriate, robust and clinically plausible" although the manufacturer identified 4 other large (>2000 patients) data sets of MS patients, but collaboration could not be established. There is

no attempt by the manufacturer to assess the generalizability of the London Ontario data to the population that they primarily focus on in the analysis (i.e. non-HA/RES RRMS patients).

The model allows other sources of disability progression to be incorporated in the model other than the 'all RRMS' London Ontario data (Table 27), such as the placebo arms of the TEMSO or TOWER trials and the 'active RRMS' patients from the London Ontario dataset (Table 28). The disability transition matrices estimated from the trial data allow for regression to lower EDSS states (Table 29).

Table 27: Transitions by EDSS state for 'all RRMS' patients from the London Ontario data.

		year i+1					
	EDSS	**0**	**1**	**2**	**3**	**4**	
year i	**0**	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	1	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	2	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	3	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	4	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	5	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	6	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	7	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	8	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	9	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	10	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•

Table 28: Transitions by EDSS state for 'active RRMS' patients from the London Ontario data.

		year i+1					
	EDSS	**0**	**1**	**2**	**3**	**4**	
year i	**0**	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	1	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	2	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	3	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	4	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	5	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	6	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	7	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	8	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	9	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•
	10	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	•

Table 29: Transitions by EDSS state from the TEMSO & TOWER placebo arms.

		year i+1					
	EDSS	**0**	**1**	**2**	**3**	**4**	
year i	**0**	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	1	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	2	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	3	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	4	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	5	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	6	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	7	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	8	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	9	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	
	10	• [redacted]	• [redacted]	• [redacted]	• [redacted]	• [redacted]	

Little information was provided about the methods used to calculate the transition matrices except to say that the "matrix was derived using the msm-package using individual data (extracted from pooled data) from the placebo arms of the two studies TEMSO and TOWER" and that the same approach was used as with the London Ontario data (Manufacturer's response, pg.37). No justification or rational was offered by the manufacturer as to why external data was preferred to the placebo arm of the trial data. The ERG used the manufacturer's base case model to estimate the time in each health state using the London Ontario data and the placebo arm of the TEMSO and TOWER trials (Figure 7). Transitions from the London Ontario data resulted in patients spending more time in the more severe health states.

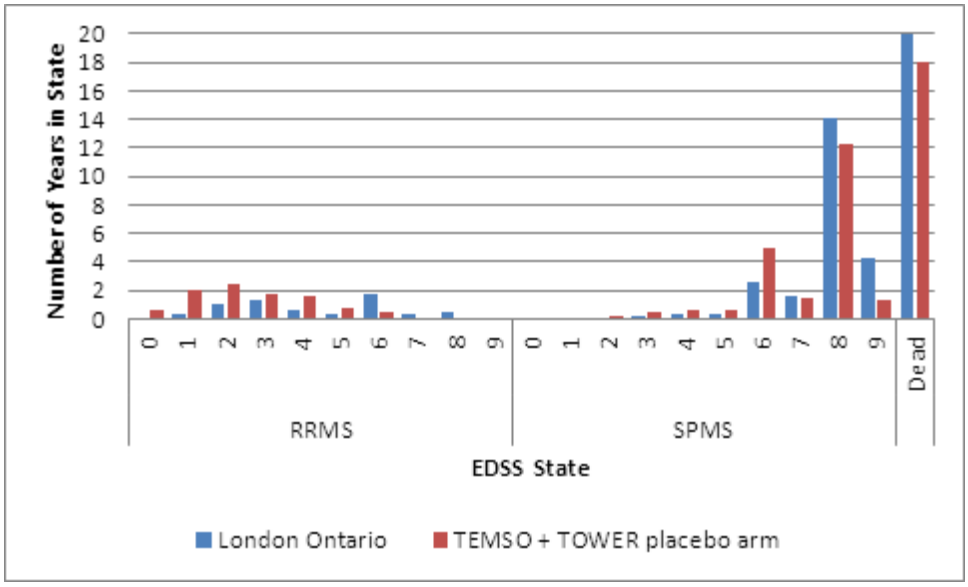


Figure 6: Average number of years in health states (ERG analysis based on manufacturer model)

The London Ontario data provided EDSS transition probabilities for all RRMS patients, all SPMS patients and a sub-group of the RRMS patients 'active RRMS' patients. Active RRMS patients were defined as "patients with ≥2 relapse events during the 2-year pre-study period, but <2 relapses in the first pre-study year". This is similar to the ABN eligibility guidelines for treatment with beta-interferon or glatiramer acetate, >2 clinically significant relapses in the previous 2 years. The manufacturer's base case analysis uses 'all RRMS' patients even though the comparators beta-interferon and glatiramer acetate are only recommended in the population more similar to 'active RRMS' patients. This suggests within the 'all RRMS' population there are also 'non-active RRMS' patients for whom beta-interferons and glatiramer acetate are not recommend and the appropriate comparator to teriflunomide would be BSC. No information has been provided on this population as the manufacturer considers teriflunomide appropriate only for patients that are currently considered appropriate for beta-interferons and glatiramer acetate. This suggests that the 'active RRMS' population from the London Ontario data may be more similar to the population of interest than the 'all RRMS' population, and therefore provide the most appropriate transition probabilities. The ERG preferred transitions calculated from the trial data because they allowed patients to improve their EDSS scores and the trial population matched the licensed population.

5.2.6.2 Conversion from RRMS to SPMS

The model presented in the submission assumed that patients face a transition probability of conversion to SPMS for each period they are in RRMS. Conversion to SPMS is dependent on the level of disability (i.e. the EDSS state). Once patients have transitioned to SPMS they do not return to RRMS. SPMS patients experience disability progression through increases in EDSS score, analogously to RRMS patients.

Conversion rates from RRMS to SPMS were also calculated from the London Ontario dataset. The median time to conversion observed in the data was used to specify an exponential distribution for EDSS 1. Then a Cox proportional hazards model was used to calculate conversion rates between EDSS 1 and all other EDSS states.

Table 30: Transition to SPMS by EDSS state

EDSS	Manufacturer's calculation of the probability of conversion to SPMS during following year
0	[redacted]
1	[redacted]
2	[redacted]
3	[redacted]
4	[redacted]
5	[redacted]
6	[redacted]
7	[redacted]
8	[redacted]
9	[redacted]

Conversion probabilities for EDSS states 0 and 9 are not calculated and are assumed in the model to take the values of 0% and 100% respectively. Table 30 shows the conversion rates by EDSS states used in the model.

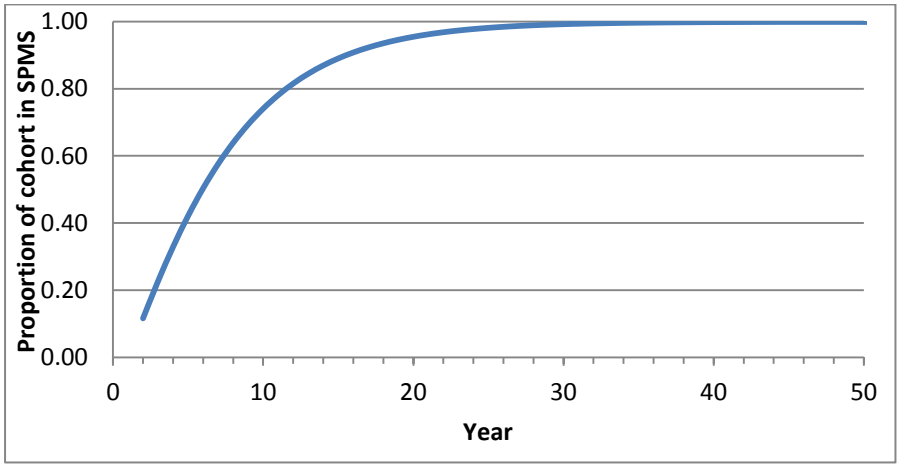


Figure 7: Conversion of BSC cohort to SPMS

Figure 8 shows the proportion of patients who have converted to SPMS states over time as predicted by the model. The model only allows for conversion in on direction i.e. from RRMS to SPMS, thus by 15 years 90% of BSC patients have converted to SPMS. The manufacturer compares this to previous data which suggests that 65% of RRMS patients not treated with DMT will develop SPMS within 15 years of diagnosis (NICE Final Scope). However, these are patients 15 years since diagnosis compared to patients 15 years since treatment initiation. Trojano et al. estimate that approximately 90% of untreated RRMS patients will transition to SPMS after 20-25 years.[33] The validity of the model will be discussed further in Section 5.2.11.

5.2.6.3 Relapse rates and severity

The relapsing-remitting nature of MS was included in the model through a probability of relapse in each cycle of the model until death. Relapse rates were modelled to depend on the EDSS state. In the model clinical progression was not affected by repeated relapses. The ERG's clinical understanding is that disability progression often occurs after relapse, such that a patient may never fully recover after experiencing a relapse (i.e. relapse initiates disability progression). The correlated nature of relapses and disability was not included in the model.

The model also considers the severity of relapses by taking into account the proportion of relapses that lead to hospitalisation. While the severity of the relapses does affect the model outcomes through costs and HRQoL it is not correlated with the disability progression.

EDSS state-dependent relapse rates split by MS classification (RRMS or SPMS) were included in the model. Relapse rates were not estimated from the London Ontario dataset, due to a lack of data, but were collected from two published sources, Patzold et al. 1982 and Held et al. 2005.[34, 35] It is not clear how these studies were chosen as no systematic review of relapse rates was described by the manufacturer.

Relapse rates calculated from Patzold et al. have been used in previous NICE appraisals (TA 127 and TA 254), however the manufacturer chose to use those from Held et al. because they are more current estimates from a dataset of more patients (821 versus 102).[34, 35] To calculate the relapse rate by EDSS, data from each source on the relapse rate per year since diagnosis was combined with data from Orme et al. on the number of patients in each EDSS state by the number of years since diagnosis.[36]

Table 31 reports the relapse rates estimated from Held et al. used in the base case model and the relapse rates estimated from Patzold et al. used in sensitivity analysis.[34, 35] These sources resulted in very different estimates of annual relapse rates (i.e. 0.312 versus 0.818 for patients with SPMS and EDSS 7-9). The values used in the sensitivity analysis are similar to those used in base case analysis of TA 254 as similar methods and data were used.

Table 31: Natural history relapse rates by EDSS score and MS classification

EDSS scale	Relapse Rate (mean number of events per person-year)			
	Base case		Sensitivity analysis	
	RRMS	SPMS	RRMS	SPMS
0	0.905	0.000	0.725	0.000
1	0.905	0.000	0.743	0.000
2	0.895	0.851	0.690	0.448
3	0.899	0.910	0.723	0.788
4	0.900	0.872	0.707	0.567
5	0.881	0.860	0.599	0.517
6	0.859	0.847	0.509	0.445
7	0.856	0.818	0.504	0.312
8	0.856	0.818	0.504	0.312
9	0.856	0.818	0.504	0.312
Source used in derivation	Held et al. 2005, [34] Orme et al. 2007[36]		Patzold et al. 1982, [35] Orme et al. 2007[36]	

In the response to the ERG's points for clarification the manufacturer reports the average number of relapses in the placebo arms of the TEMSO and TOWER trials is [redacted] in the first year compared to the BSC modelled estimate of [redacted] in the first year (Response to Points for Clarification, Table B10). No justification was given for using external literature rather than the results from the placebo arms of the trials. The large absolute difference between the model predicted and trial observed values is discussed further in Section 5.2.13.

The model distinguishes relapses by those that lead to hospitalisation or not. The manufacturer subsequently applied different costs and HRQoL estimates to relapses according to whether patients were assumed to be hospitalised or not. Two sources of natural history of the proportion of relapses leading to hospitalisation were tested in the model, Dee et al. 2012 and the placebo arms of the TEMSO and TOWER trials.[37] Dee et al. estimate that 80% of relapses lead to hospitalisation compared to 69.3% in the TEMSO and TOWER trials. There was assumed to be no difference in hospitalisation between EDSS states, MS classifications or treatments. The base case considers Dee et al. as the manufacturer considers this source more representative of the UK population, but the trial data is considered in a sensitivity analysis.

5.2.6.4 Mortality

All-cause mortality is applied to the model taking into account the proportion of females being treated in the population and the age at each cycle in the model. The additional mortality from MS is applied as a multiplier dependant on the EDSS state. It is assumed that there is no excess mortality risk for patients in EDSS 0. Similar methods have been used in previous NICE submissions (TA 127 and TA 254). In the model there is no difference in mortality rates between RRMS and SPMS patients,

although SPMS patients experience disability at a faster rate than an RRMS patient thus indirectly increasing mortality risk.

All-cause mortality rates for the general population were derived from age and gender specific mortality rates from the Office of National Statistics 2008-2010 data for England and Wales. A systematic review of mortality by EDSS was undertaken by the manufacturer, but they reported that no additional literature to inform mortality predictions was found compared to the methods used in previous NICE appraisals (TA 127 and TA 254). In common with the models used to inform other Nice technology appraisals in MS, the probabilities were adjusted for the MS population using the mortality ratios by mild, moderate and severe disability categories reported by Pokorski 1997 (Table 32).[1] The Pokorski study was a long-term survival analysis of 2348 patients with MS in Denmark from 1951 to 1990. It was assumed that mortality does not differ between RRMS and SPMS patients.

Table 32: Mortality multipliers by EDSS state

EDSS 0	EDSS 1	EDSS 2	EDSS 3	EDSS 4	EDSS 5	EDSS 6	EDSS 7	EDSS 8	EDSS 9
1	1.432	1.600	1.637	1.674	1.842	2.273	3.097	4.447	6.454

5.2.7 Treatment effectiveness and extrapolation

Treatment effectiveness is captured in the model by applying separate hazard ratio estimates to the natural history estimates for both disease progression and the rate and severity of relapse. The model uses treatment effects that have been estimated from the mixed treatment comparison previously described. The manufacturer assumed that the hazard ratios remain constant while remaining on treatment, throughout the 50 year time horizon of the model. The model assumes that there is no remaining treatment effect once DMT is stopped. The time horizon of 50 years is longer than what was applied in TA 32 and TA127 but the same as was used in TA 254. However, in TA 254 treatment waning was an issue consider in scenario analyses by the ERG to reflect the uncertainty about projecting over such a long time period. The manufacturer in this case also explored the waning of treatment efficacy in scenario analyses.

This section describes how the treatment effects are applied in the model, assessing the impact of treatment on each element of the model (disability progression, conversion, relapse and mortality).

5.2.7.1 Disability Progression

The effect of treatment on disease progression is modelled as a hazard ratio of confirmed disability progression (3 month sustained accumulation of disability). The model applies hazard ratios for each modelled treatment vs. BSC to the individual natural history transitions for RRMS.

The hazard ratios are derived from the MTC. As discussed previously the results of the MTC are not robust given the exclusion of a large number of studies and the heterogeneity across those studies included. As discussed previously, the MTC predicts teriflunomide is more effective (on average) at slowing progression of disability than all commonly used beta-interferons and glatiramer acetate. The main concerns with the results are,

i) The results for IFN-β 1b are worse than BSC which raises issues about the validity of the MTC.
ii) The estimate versus Rebif 44µg from the MTC appears more favourable towards teriflunomide compared to the direct head to head evidence.
iii) Using the blended comparator means the effect size applied for beta-interferons is more favourable towards teriflunomide compared to the individual estimates of 3 of the components of the blend.

Further exploration around these estimates will be discussed in Section 6.

The hazard ratios for disease progression are assumed to be constant (time independent) and are only applied while on treatment. All patients are modelled to follow a natural history path (i.e. BSC) once they fall outside the RRMS treatment range or have converted to SPMS.

Table 33 summarises the hazard ratios of progression used in the model as calculated in the MTC. The blended comparator which is used as the main comparator in the manufacturer's analysis is calculated as the market share weighted average of Rebif 22µg, Rebif 44µg, Avonex 30µg, Betaferon 250µg and Copaxone. From the table we see that Betaferon 250µg is estimated to be less effective than BSC.

Table 33: Hazard ratios of progression

Comparator	*Disease progression*
Natalizumab	**0.58**
Teriflunomide	**0.71**
IFN-β 1a (Rebif 44µg)	**0.79**
IFN-β 1a (Avonex 30µg)	**0.91**
IFN-β 1b	**1.21**
Glatiramer acetate	**0.93**
Fingolimod	**0.75**
IFN-β 1a (Rebif 22µg)	[redacted]
BSC	**1.00**
Blended Comparator	**0.91**

5.2.7.2 Conversion from RRMS to SPMS

The model does not apply a treatment effect to the conversion rate from RRMS to SPMS. Conversion rates are however dependent on EDSS states (Table 30). It is therefore important to note that by modifying progression through these EDSS states, the treatments also indirectly modify the proportions of the cohort converting from RRMS to SPMS over the modelled time horizon. Figure 9 below shows the model predicted proportion of the population in the SPMS state over time for the two treatments showing a small but noticeable delay in conversion for patients treated with teriflunomide as compared to those treated with the blended comparator. In the later years teriflunomide patients are more likely to be SPMS as patients in the blended comparator arm are more likely to be dead (Figure 11).

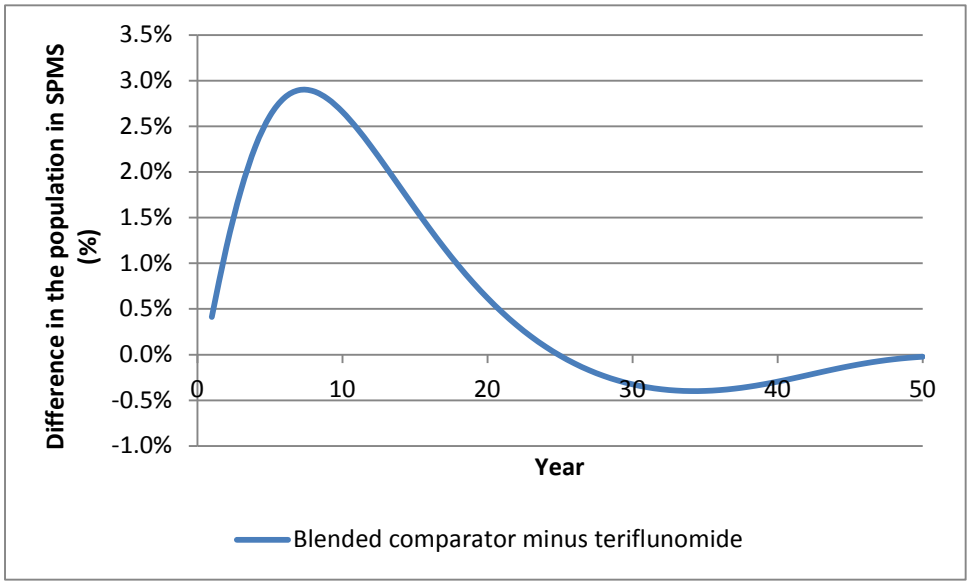


Figure 8: Treatment effect on SPMS population over time (ERG analysis based on the manufacturer model)

5.2.7.3 Relapse rate and severity

The effect of treatment on the mean number of relapses is modelled as a relative rate. Table 34 below summarises the relative risk used in the model. As with the hazard ratios for progression, the relative rates for relapse are derived from the MTC. The relative rates for relapse are assumed to be constant and only applied while on treatment. All patients are modelled to follow a natural history path (i.e. BSC) once they fall outside the RRMS treatment range or have converted to SPMS.

Table 34: Relative rate of relapse

Comparator	Relative annualised rate of relapse
Natalizumab	0.31
Teriflunomide	0.67
IFN-β 1a (Rebif 44µg)	0.62
IFN-β 1a (Avonex 30µg)	0.78
IFN-β 1b	0.68
Glatiramer acetate	0.64
Fingolimod	0.46
IFN-β 1a (Rebif 22µg)	[redacted]
BSC	1.00
Blended Comparator	0.68

Figure 10 below shows the impact of applying the natural history relapse rates in Table 31 with the treatment effects in Table 34. The graph shows that on average a patient on teriflunomide is predicted to initially experience more relapses resulting in hospitalisation but fewer relapses not resulting in hospitalisation. However after approximately 10 years patients on teriflunomide experience more of both types of relapses. This occurs because patients are more likely to have a relapse in lower stages of EDSS progression; the highest rate of relapse is 0.91 for EDSS scores 0-1 of RRMS and the lowest is 0.82 for EDSS scores 7-9 of SPMS. Since teriflunomide slows transitions through EDSS teriflunomide patients experience more relapses in the model.

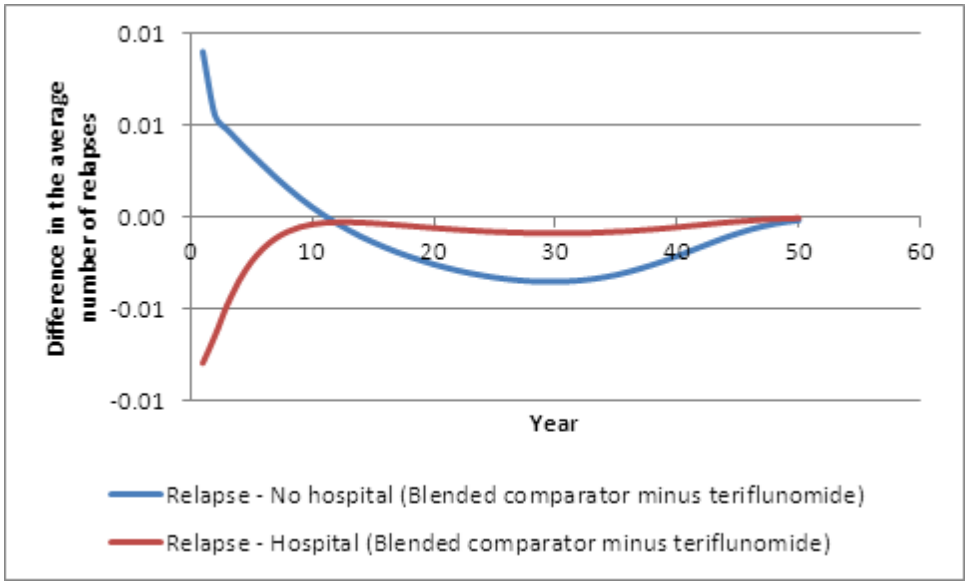


Figure 9: Cumulative relapse (ERG analysis based on manufacturer model)

The manufacturer found limited published literature on the treatment effect of reducing the proportion of relapses which lead to hospitalisation. Data from the TEMSO trial was used to estimate the relative

risk of teriflunomide (Table 35). Data from the TRANSFORMS trial was used to estimate the relative risk of Avonex 30µg and Fingolimod. The other beta-interferons and glatiramer acetate were assumed to have the same relative risk as Avonex 30µg and natalizumab was assumed to have the same as fingolimod.

Table 35: Relative risk of relapse leading to hospitalisation

Comparator	*Risk of relapse hospitalisation*
Natalizumab	0.60
Teriflunomide	0.57
IFN-β 1a (Rebif 44µg)	0.49
IFN-β 1a (Avonex 30µg)	0.49
IFN-β 1b	0.49
Glatiramer acetate	0.49
Fingolimod	0.60
IFN-β 1a (Rebif 22µg)	0.49
BSC	1.00
Blended Comparator	0.49

Any treatment that impacts on progression will also have an indirect impact on relapse. This is in addition to any direct impact the treatment has on relapse effect. The correlation between the direct and indirect treatment effects of relapse has not been captured in the model and potentially results in double counting of treatment effect. We will return to exploring the impact of this structural weakness in the manufacturer's model in Section 6.

5.2.7.4 Discontinuation

Patients were discontinued from treatment in the model for the following reasons,

- When patients convert from RRMS to SPMS
- When patients' disease progression reaches an EDSS score greater than 6. This upper bound for the provision of treatment is consistent with the guidelines of the Association of British Neurologists (ABN 2009).
- Treatment specific discontinuation derived from the mixed treatment comparison.

The discontinuations are applied in the model throughout the on-treatment period. The weighted average 2 year probability of discontinuation on teriflunomide 14mg is estimated from the TEMSO, TOWER, TENERE and phase II trials, 28.7%. This weighted average 2 year probability is then converted to odds and used within the MTC. Results of the MTC are then used to calculate 1 year probabilities, [redacted] in the case of teriflunomide (Table 36). The 1 year probability of discontinuation is

assumed for the first two years of the model and then discontinuation is assumed to decrease by 50% due to fewer patients having adverse events. A 75% decrease in the 1 year probability of discontinuation is tested in the sensitivity analysis.

The manufacturer assumes that discontinuation will decrease after 2 years because after 2 years patients will have adjusted to treatment use and those who do not tolerate treatment will have discontinued. This assumption is not supported by any evidence and the manufacturer do not consider the alternate possibility that discontinuation increases because of loss of response or poor compliance which may influence longer term discontinuation and may not be captured by these short time horizon trials.

Table 36: Annual all-cause withdrawal

Comparator	*Annual probability of withdrawal*
Natalizumab	[redacted]
Teriflunomide	[redacted]
IFN-β 1a (Rebif 44µg)	[redacted]
IFN-β 1a (Avonex 30µg)	[redacted]
IFN-β 1b	[redacted]
Glatiramer acetate	[redacted]
Fingolimod	[redacted]
IFN-β 1a (Rebif 22µg)	[redacted]
BSC	[redacted]
Blended Comparator	[redacted]

In the model the probability of withdrawal has the potential to have an important and perverse effect on the ICER. This occurs because existing DMTs have ICERs which exceed NICE's usual threshold and hence have been funded through the Risk Sharing Scheme. However, it is clearly plausible, but also perverse, when applying NICE's usual threshold range to show that treatments become more cost-effective (i.e. the ICER decreases) the higher the rate of withdrawal. In the base case model patients that withdrawal are treated with BSC. The higher withdrawal rate results in patients being treated with BSC more quickly. Since active treatments are not cost-effective compared to BSC the quicker patients on active treatment are switched BSC the lower the ICER compared to another active treatment.

5.2.7.5 Mortality

No direct treatment effect on mortality is modelled. However, as treatments affect EDSS state progression and EDSS states affect mortality there is an indirect treatment effect on mortality in the model. This difference in mortality is depicted in Figure 11.

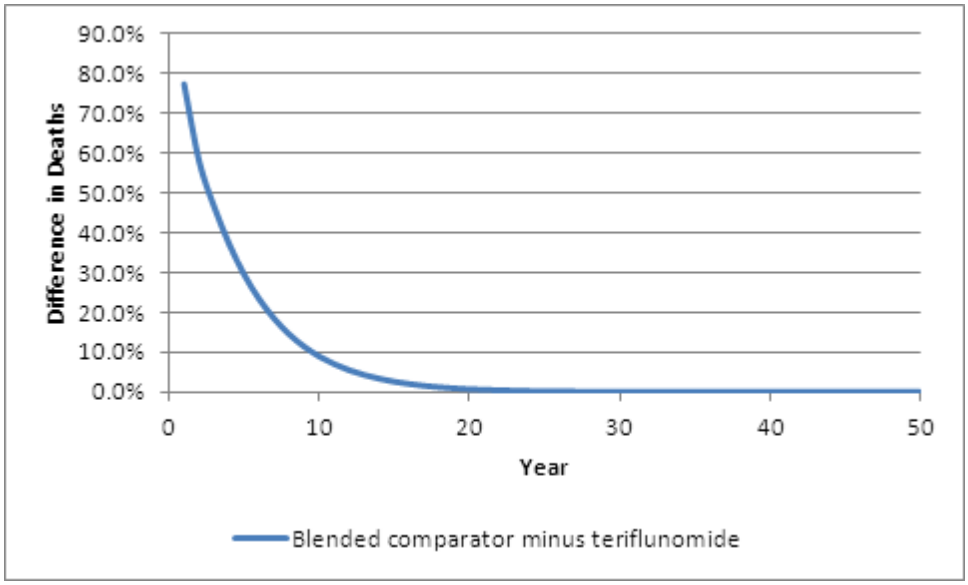


Figure 10: Incremental mortality (ERG analysis based on the manufacturer model)

5.2.7.6 Adverse events

Adverse events were included in the model if there was a probability of occurrence of ≥4% compared to placebo. According to the manufacturer the reporting of trial data was variable, as they did not want to bias the results against those treatments that had more detailed publications only common (i.e. ≥4%) adverse events were included. This resulted in some severe adverse events such as cardiac issues associated with fingolimod and progressive multifocal leukoencephalopathy associated with natlizumab not being included in the model. The manufacturer suggests that by only including common adverse events the cost and disutility of DMTs may be slightly underestimated. The actual effect of the omission of serious events is unknown.

5.2.7.7 General issues concerning treatment effectiveness

The blended comparator uses an average hazard ratio of progression, risk of relapse, risk of hospitalisation, withdrawal, adverse events and costs for all treatments but BSC, fingolimod and natalizumab. As discussed previously a blended comparator does not meet the NICE methods guide reference case. However if justification for a blended comparator was found it should not be calculated as the outcomes from the mean relative risks of treatments, but should be calculated as the mean outcomes from the treatments. Given the correlation between costs and benefits in the model these two methods may differ significantly and will be explored further in Section 6.

Treatment effects on progression, risk of relapse and withdrawal are derived from the MTC. The treatment effects estimated from the MTC differ from those calculated in the head-to-head trial (TENERE). Further exploration around these estimates will be discussed in Section 6.

5.2.8 Health related quality of life

The cost-effectiveness model assessed the health-related quality of life (HRQoL) impact of the different treatments for both the patient and their carer, and incorporated these in the assessment of quality-adjusted life years (QALYs). For the patients, estimates of HRQoL were used to account for differences in disability progression (EDSS), conversion to SPMS, relapses and adverse events. The HRQoL impact on the carer was dependent on the EDSS state of the patient.

5.2.8.1 HRQoL of Disability States

A systematic review of HRQoL for patients with RRMS was undertaken by the manufacturer to obtain all relevant quality of life studies in MS. The manufacturer identified two relevant sources of HRQoL by disability state, Orme et al. 2007 and Gani et al. 2008.[36, 38] In the base case, HRQoL were based on responses from the 2005 UK MS Survey (Table 37).[36] This survey has been criticised in previous NICE evaluations (TA 127) because of the low response rates, selection bias, unrepresentative population and self-reported severity. The HRQoL from Orme et al. is preferred by the manufacturer and used in the base case. The manufacturer considered the Orme study to be comparable to other studies and to relate well to the population of interest as it is a survey of UK patients.

Table 37: Utilities by EDSS state used in the base case model

EDSS	0	1	2	3	4	5	6	7	8	9
RRMS	0.870	0.799	0.705	0.574	0.610	0.518	0.460	0.297	–0.049	–0.195
SPMS	0.825	0.754	0.66	0.529	0.565	0.473	0.415	0.252	–0.094	–0.240

The TEMSO trial also collected HRQoL using EQ-5D. The EQ-5D data from the placebo arm of the TEMSO trial was included to test the sensitivity of the model. As TEMSO did not collect data on patients in EDSS states greater than 6, the HRQoL for these states is from Orme et al. The utility decrement for SPMS was also taken from Orme et al. as the manufacturer considered the SPMS patient numbers in TEMSO to be too low for robust analysis of the HRQoL data. The HRQoL from TEMSO was higher in all EDSS states than the estimates from Orme et al. (Figure 12).

[redacted]

[redacted]11: Comparison of HRQoL between Orme et al. and TEMSO (Manufacturer's submission, Figure B7.4.5)

The manufacturer considered the HRQoL from TEMSO to be less representative of the UK population as it is an international study. The ERG was not concerned that health states were derived from an international population since UK specific preferences were applied to these health states. The ERG considered the TEMSO data more applicable to the treatment population because it is more reflective of first-line patients that are likely to receive teriflunomide.

Figure 12 also demonstrates the steep decline in HRQoL in the later stages of progression. With EDSS scores of 8 or 9 patients have negative HRQoL suggesting that patients are worse off than if they were dead. The effect of these negative HRQoL is explored further in Section 5.3. The manufacturer does not consider HRQoL data used in previous NICE technology appraisals, nor do they consider applying HRQoL decrements from Orme to the TEMSO trial data rather than using the absolute HRQoL values (Figure 13).

[redacted]

[redacted] 12 [redacted] The ERG explores further HRQoL data and methods for combining this data in the model in Section 6.2.6.

5.2.8.2 HRQoL of Relapse

The disutility of relapse is based on data from the UK MS Study (Orme et al. 2007).[36] However, no UK studies are identified which report HRQoL by relapse severity. Data from a US study was combined with data from Orme et al. to estimate disutility of relapses leading to hospitalisation. Prosser et al. 2003 used a standard gamble to value hypothetical health states in MS; data was collected on 62 patients diagnosed with RRMS and 67 members of the general population in Boston Massachusetts in 1999.[39] HRQoL for relapse was not reported in this study but was reported in a subsequent study by the author citing the 2003 study. In Prosser et al. 2004 the authors report mild to moderate relapse results in a HRQoL of -0.091 and severe relapse results in a HRQoL decrement of -0.302.[40] The manufacturer used the difference in the HRQoL for the mild/moderate relapse and the severe relapse and applied this difference to the UK disutility of relapse of -0.071 from Orme et al. 2007. To do this the manufacturer assumed the disutility from Orme et al. applied to mild/moderate patients (i.e. patients that were not hospitalised). The manufacturer calculated a disutility of relapse for hospitalised patients of -0.2356. In previous analyses all relapses (both hospitalised and not) were assumed to have a disutility of –0.071 (TA 254). It is not clear that the manufacturer's assumption that the disutility of -0.071 only applies to mild/moderate patients is correct. Further the manufacturer has used international (i.e. US) valuations of health states contrary to their previous argument that TEMSO should not be used because it collects health states from an international population. The manufacturer demonstrated in their one-way sensitivity analysis (Manufacturer's submission, pg. 317) that changing the disutility on relapses leading to hospitalisation by plus or minus 10% changes the ICER by less than £30. The ERG tests the effects of relapse on the ICER in Section 5.3.3.

5.2.8.3 HRQoL of Adverse Events

The total disutility of an adverse event is calculated by multiplying the disutility by the duration that the event is likely to occur. Several assumptions have been made regarding the duration of adverse events. Most studies were conducted in the UK population [41-43]; however, those utilities using Sullivan et al. 2006 were from a population in the US.[44] Studies used either standard gamble methods or EQ-5D. Values were taken from a mix of populations: from patients or from the general population using hypothetical health states. The sources and estimates of disutility are reported in Table 38. The manufacturer demonstrates in their sensitivity analysis that changing the adverse events disutility rates by plus or minus 10% changes the total QALYs in the model by plus or minus 0.002 QALYs.

Table 38: Utilities and durations of adverse events

	Utility of event	Source of utility of event	Utility assumptions	Duration of event	Source for duration of event	Annual disutility
Teriflunomide						
Nausea	-0.048	[42]	Standard gamble of UK general population	1 day	Assumption	-0.0001
Diarrhoea	-0.047	[42]	Standard gamble of UK general population	3 days	Assumption	-0.0004
Hair thinning	-0.114	[42]	Standard gamble of UK general population	1 year	[45]	-0.1140
Natalizumab						
Fatigue	-0.073	[42]	Standard gamble of UK general population	1 week	Assumption	-0.0014
Arthralgia	-0.040	[44]	EQ-5D of US general population. Based on ICD-9 719 Joint Disorder	1 month	Assumption	-0.0034
Headache	-0.030	[44]	EQ-5D of US general population. Based on ICD-9 346 Migraine	2 days	Assumption	-0.0002
Glatiramer acetate						
Injection site reaction	-0.011	[41]	Standard gamble of Scottish population	1 day every dose (every day [Copaxone SmPC])	Assumption	-0.0110
Immediate postinjection systemic reactions	-0.041	[44]	EQ-5D of US general population. Based on ICD-9 413 Angina Pectoris	0.5 days	Assumption	-0.0001
Fingolimod						
Lower respiratory tract infection	-0.020	[43]	Assumed same as disutility value for influenza like symptoms	2 weeks	[46]	-0.0008
Back pain	-0.046	[43]	EQ-5D of US general population. Based on ICD-9 724 Back Disorder	2 weeks	Assumption	0.0018
Diarrhoea	-0.047	[42]	Standard gamble of UK general population	3 days	Assumption	-0.0004
IFNβ-1a (Avonex 30µg)						
Headache	-0.030	[44]	EQ-5D of US general population. Based on ICD-9 346 Migraine	2 days	Assumption	-0.0002
Influenza-like symptoms	-0.08	[43]	EQ-5D on English patients with H1N1 virus	1 day every dose (every other day [Avonex 30µg SmPC])	[43]	-0.0114
IFNβ-1a (Rebif)						
Headache	-0.030	[44]	EQ-5D of US general population. Based on ICD-9 346 Migraine	2 days	Assumption	-0.0002
Influenza-like symptoms	-0.08	[43]	EQ-5D on English patients with H1N1 virus	1 day every dose (thrice weekly [Rebif SmPC])	[43]	-0.0343
Injection site reaction	-0.011	[41]	Standard gamble of Scottish population	1 day every dose [thrice	Assumption	-0.0047

	Utility of event	Source of utility of event	Utility assumptions	Duration of event	Source for duration of event	Annual disutility
				weekly [Rebif SmPC])		
IFNβ-1b						
Influenza-like symptoms	-0.08	[43]	EQ-5D on English patients with H1N1 virus	1 day every dose (every other day [Betaferon SmPC])	[43]	-0.040
Injection site reaction	-0.011	[41]	Standard gamble of Scottish population	1 day every dose (every other day [Betaferon SmPC])	Assumption	-0.0055

5.2.8.4 HRQoL of Carers

HRQoL for carers has been included in the model using the method developed by Gani et al. [38] and used in previous NICE evaluations (TA 254). It was assumed that disutility of caregivers had a maximum value of 0.14, based on a value accepted by NICE in an assessment of treatments for Alzheimer's disease. The value 0.14 was divided by the amount of time caregiving in EDSS state 8.5-9.5 from the UK MS Survey to estimate the ratio: disutility/time caregiving. Estimates for other EDSS states were calculated by multiplying this ratio by the amount of time caregiving in each EDSS state. These disutilities are reported in Table 39.

Table 39: Carer disutility used in the model

EDSS state	0	1	2	3	4	5	6	7	8	9
Disutility	0.000	-0.001	-0.003	-0.009	-0.009	-0.020	-0.027	-0.053	-0.107	-0.140

5.2.9 Resources and costs

The manufacturer stated that costs used in the model were conducted from an NHS and PSS perspective and only considered resources relevant to the management of MS. Although this is the stated perspective there is uncertainty surrounding what costs were included in some of the sources used, particularly the direct non-medical costs. The manufacturer conducted a systematic review to obtain all relevant cost and resource use studies in MS. The manufacturer separated the analysis of costs into, disease costs by disability and relapse, acquisition costs, administration costs, monitoring costs and costs of adverse events. Most costs seem to be reported in 2012 GBP except the acquisition costs.

5.2.9.1 Disease costs

The progressive disability of MS is assumed to result in increasing costs. Two sources of costs were considered for disability health states, each dividing costs by EDSS scores. In the base case the manufacturer used costs from Tyas et al. 2007 [47], because "it is well cited and consistent with two previous NICE submissions". These costs included the direct medical costs and direct non-medical

costs funded by the government collected using a survey of 2,508 respondents.[47] It is not clear what costs are included as direct non-medical costs. In Tyas et al. no description of what constituted direct non-medical costs was provided. While some direct non-medical costs may be considered under the PSS perspective there may be other costs that do not meet this perspective. It is not known whether the direct non-medical costs meet the NICE reference case perspective. Other sources were also considered, see Figure 14. This figure demonstrates that costs do not increase linearly with increasing EDSS but there is a marked increase once mobility is affected (i.e. EDSS 6). Karampampa et al. reported other direct costs much higher than Tyas et al., this may be because informal care costs were included factoring in the productivity losses of the working caregivers.[48] These costs do not meet NICE reference case perspective.

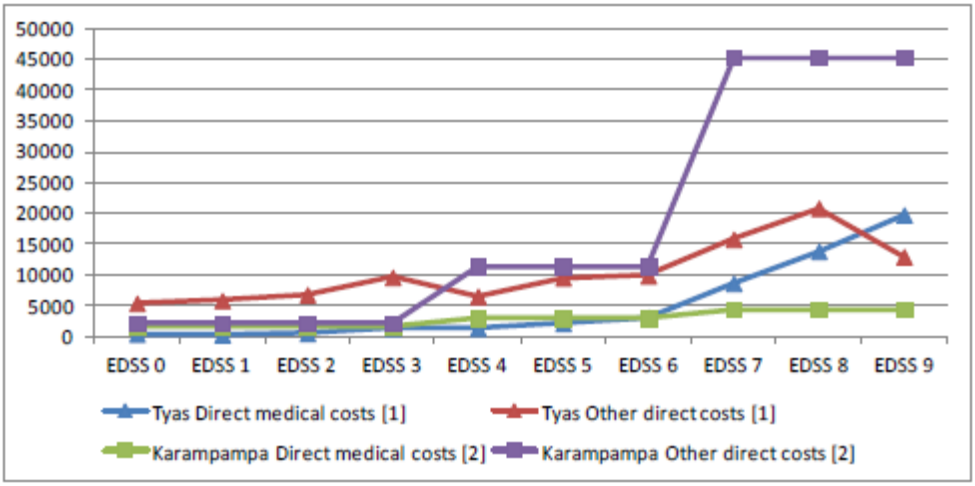


Figure 13: The natural history costs of MS compared (Manufacturer's submission, Figure B7.5.4)

The model differentiates the severity of relapses by relapses leading to hospitalisation and those that do not. The resource use of having had a relapse leading to hospitalisation was estimated from Irish neurology centres to obtain a length of hospital stay of 10.7 days; non-hospitalisation relapses were assumed to require 5 days of steroid treatment as day cases (Dee et al. 2012).[37] UK NHS reference costs were used with the resource use to estimate costs of relapse by whether or not the patient was hospitalised. Estimated costs are reported in Table 40. Costs estimated from Tyas et al. were used in the sensitivity analysis, however these costs do no differentiate by hospitalisation.[47]

Table 40: Costs associated with relapse in the economic model

Category	Base case cost [Dee, 2012]	Sensitivity analysis cost[47]
Relapse not leading to hospitalisation	£844.65	£2,006.03
Relapse leading to hospitalisation	£6,164.46	£2,006.03

The cost of relapse in previous NICE appraisals has varied. TA 254 used costs of relapse based on the 2010-2011 Nation Tariff of "admitted patient care & outpatient procedure tariff, AA30Z multiple

sclerosis non-elective tariff" of £3,039. The 2012-2013 Nation Tariff (Department of Health, 2012) is £3,739. TA 127 used cost of relapse of £228. A wide range of costs for relapse have been used and could be tested in the sensitivity analysis. The manufacturer only varied costs of relapse by plus or minus 10%. This small change in costs resulted in very small differences in overall costs, but this may not capture the full range of the uncertainty. The ERG explores the sensitivity around relapse further in Section 5.3.3.

5.2.9.2 Drug acquisition costs

Many of the disease modifying treatments are made available through the MS Risk Sharing Scheme. The acquisition costs of beta-interferon and glatiramer acetate used in the model are those set by the MS Risk Sharing Scheme and published by the UK Department of Health in 2002.[2] The prices in the model have not been updated for inflation. The prices of fingolimod and natalizumab are the list prices publicly available from the BNF.[49] Fingolimod was approved under a patient access scheme (PAS), but as the PAS costs are not public it has not been incorporated into this analysis, however a range of fingolimod prices (£11,000 and £13,000) were tested in scenario analysis. The acquisition costs used in the model are detailed in Table 41. The cost of the blended comparator used in the main analysis of the model is £7,712. This is calculated as the market share weighted average of the acquisition costs of the beta-interferons and glatiramer acetate.

Table 41: Acquisition costs (Manufacturer's submission, Table B7.5.1)

Treatment	Dosage	Treatment Regimen	Total Annual Cost
Teriflunomide	14 mg	Once daily oral form	£13,529
Teriflunomide (PAS)	14 mg	Once daily oral form	[redacted]
IFNβ-1a (Rebif)	44µg	Self-injected subcutaneously	£8,942
IFNβ-1a (Rebif)	22µg	Self-injected subcutaneously	£7,513
IFNβ-1a (Avonex 30µg)	30 µg	Self-injected subcutaneously	£8,502
IFNβ-1b	250 µg	Self-injected subcutaneously	£7,259
Glatiramer acetate	1 mL vial 20 mg/mL	Self-injected subcutaneously	£5,823
Fingolimod	28-tab pack 0.5 mg capsule	Once daily in oral form	£19,110
Natalizumab	15 mL vial 20 mg/mL	Intravenous injection by a health profession every 4 weeks	£14,690

5.2.9.3 Administration costs

The manufacturer assumes that there are no administration costs for teriflunomide since it is an oral treatment. All other treatments were assumed to have administration costs as described in Table 42. Administration costs of beta-interferon and glatiramer acetate were assumed to be the one off cost of training patients to self-administer. The value of £174 used in the model is more than double the cost of training self-administration of £78 used in TA 254 for beta-interferons. It is unclear from the ERG report of TA 254 what might be causing these differences, however they may be due to the expected

length of time of training or the type of specialist used for training. The cost of administering fingolimod in TA 254 is said to be £0 as it is an oral treatment. The cost of administering natalizumab in TA 127 is £1,062 per year, based on a half day visit to a neurology clinic/ward. The administration costs may bias the results in favour of teriflunomide given that previous sources used lower administration costs for its comparators however the ERG found the model was not sensitive to these differences.

Table 42: Administration costs (Manufacturer's submission, Table B7.5.2)

Treatment	Year 1 cost	Year 2 cost	Source	Justifiction
Teriflunomide	£0	£0	-	Oral so no administration costs
Beta-interferons and glatiramer acetate	£174	£0	PSSRU 2012 Clinical nurse specialist with qualifications per hour (3 hours @ £58)	3 hours required for nurse to teach self-administration, no administration following initial training
Fingolimod	£474	£0	NHS Reference Costs AA30C Medical Care of Patients with Multiple Sclerosis without CC Day Case @ £474 per visit	Continuous ECG and blood pressure monitoring for 6 hours following first dose
Natalizumab	£6162	£6162	NHS Reference Costs AA30C Medical Care of Patients with Multiple Sclerosis without CC Day Case @ £474 per visit	13 infusions per year

5.2.9.4 Monitoring costs

The resource use and associated unit costs are described in Tables B7.5.3 and B7.5.4 of the manufacturer's submission. The references for resource use were obtained from the SmPCs of each product where possible, and using the NICE fingolimod costing template. The ERG considered this a reasonable approach and on further exploration found the model was not sensitive to changes in the monitoring costs.

5.2.9.5 Adverse event costs

The adverse events and associated costs used in the model are provided in Table 43 below. For many adverse events the associated cost is assumed to be a 6 minute telephone call between the patient and their MS nurse, this includes headaches, hair thinning, nausea, injection-site reaction, influenza-like symptoms, immediate post injection systemic reaction and fatigue. Other adverse events, such as diarrhoea and lower respiratory tract infection assume a GP consultation plus treatment. No evidence was presented on the resource use associated with adverse events. It is not evident whether patients will only use a telephone call to an MS nurse or whether they would be more likely to visit their GP. The ERG considered the low costs of these adverse events might be conservative with respect to the cost-effectiveness of teriflunomide versus the blended comparator, as the probability of these low cost events is higher for the blended comparator than for teriflunomide. However, further exploratory analysis showed that including a GP visit instead of phone call decreased the ICER by less £400.

Table 43: Cost of treatment adverse events

Adverse Event	Resource Use	Source	Source Cost	Unit Cost used in Model
Headache	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Diarrhoea	GP consultation + loperamide	PSSRU 2012 GP with qualifications surgery consultation cost per patient contact last 11.7 min. BNF April 2013 loperamide hydrochloride 2 mg capsules, 30-cap pack	£43 £0.98	£43.98
Hair thinning	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Nausea	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Lower respiratory tract infection	GP consultation + 5-day regimen of amoxicillin	PSSRU 2012 GP with qualifications surgery consultation cost per patient contact last 11.7 min. BNF April 2013 amoxicillin (as trihydrate) capsules 250 mg, net price 21	£43 £0.95	£86.95
Back pain	GP consultation	PSSRU 2012 GP with qualifications surgery consultation cost per patient contact last 11.7 min.	£43	£43.00
Injection-site reaction	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Influenza-like symptoms	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Immediate post injection reaction	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Fatigue	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80
Arthralgia	MS nurse phone call	PSSRU 2012 Clinical nurse specialist with qualifications and average length of telephone call	£58 per hour 6 minutes	£5.80

5.2.10 Cost effectiveness results

The manufacturer reports both deterministic and probabilistic results. Here we present the both for the manufacturer's original model and the corrected model. With the response to the points for clarification the manufacturer presented a corrected model after discovering an error in the calculation for the duration of injection-site reactions, but this only affected the blended comparator. The corrected model results in a lower ICER for both deterministic (Table 44) and probabilistic analyses (Table 45).

Table 44: Deterministic results comparing the manufacturer's original base case to the corrected base case

Technology	Manufacturer's original base case			Manufacturer's corrected base case		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 45: Probabilistic results comparing the manufacturer's original base case to the corrected base case

Technology	Manufacturer's original base case			Manufacturer's corrected base case		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The manufacturer also presents the fully incremental analysis, which is more pertinent to NICE decision making. The deterministic (Table 46) and probabilistic (Table 47) results are presented for both the manufacturer's original and corrected model. As expected there are no differences in the deterministic results and the variation in the probabilistic results is due to the uncertainty in the estimates.

Table 46: Deterministic results comparing the manufacturer's original base case to the corrected base case in a fully incremental analysis

Technology	Manufacturer's original base case			Manufacturer's corrected base case		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Glatiramer acetate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1b	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 47: Probabilistic results comparing the manufacturer's original base case to the corrected base case in a fully incremental analysis

Technology	Manufacturer's original base case			Manufacturer's corrected base case		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Glatiramer acetate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1b	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The ICER of teriflunomide in the original fully incremental analysis was £[redacted] per QALY gained, as reported by the manufacturer compared to £[redacted] per QALY gained in the corresponding deterministic analysis. The deterministic and probabilistic model results were considerably different.

The average total QALYs range from [redacted], the QALY results are very low given that the time horizon of the analysis is 50 years. This result calls into question the face validity of the model. This issue is explored further in Section 5.3.2.

The addition of the patient access scheme price is considered in Section 5.2.11.1.

5.2.11 Sensitivity analysis

The manufacturer undertook multiple structural and sensitivity analyses. Clinical data was explored at the 95% confidence intervals or +/-10% using one-way sensitivity analysis. The most sensitive inputs were the hazard ratios on progression and the withdrawal rates. The results were also highly sensitive to the trials used in the MTC. When the pre-2000 trials were included in the MTC the manufacturer's original deterministic ICER increased from £[redacted] to £[redacted] per QALY gained. The blended comparator was much more effective when the pre-2000 publications were included in the MTC because the hazard ratio on progression of Betaferon was no longer greater than one. The ICER was at its lowest when the natural history costs from Karampampa were used in the model, decreasing the manufacturer's original deterministic ICER to £[redacted] per QALY gained. The costs from Karampampa were much higher because they included caregiver productivity, and thus small effects in slowing progression resulted in larger total cost savings.

A scatter plot of the probabilistic analysis from the corrected model shows a wide variation in the simulated points (Figure 15). The probability of teriflunomide being cost-effective compared to the blended comparator at a willingness-to-pay of £20,000 per QALY is 38%.

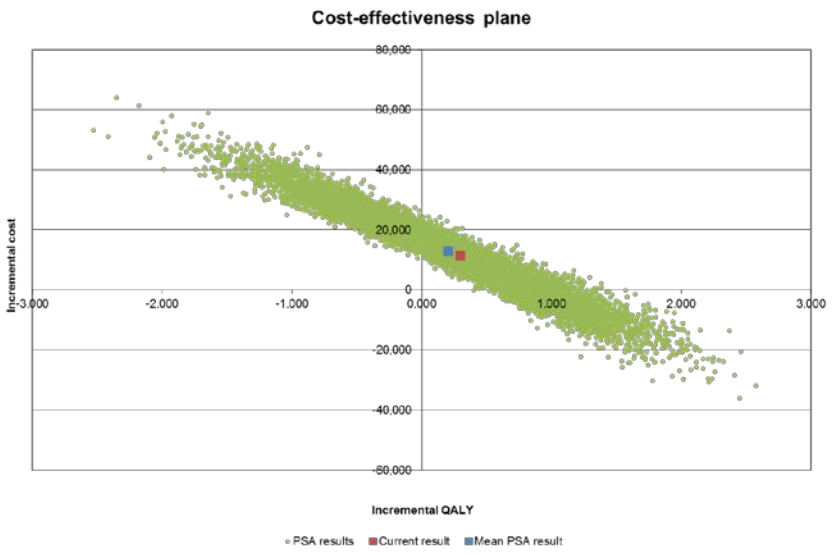


Figure 14: Probabilistic scatter plot of teriflunomide versus the blended comparator (Manufacturer's submission, Figure B7.7.3)

In addition, probabilistic sensitivity analysis has been conducted on the incremental cost-effectiveness results of teriflunomide versus all beta-interferons and glatiramer acetate (fully incremental analysis). The incremental probabilistic results are presented in Figure 16. This figure shows that teriflunomide is the least likely treatment to be cost-effective at all values of willingness-to-pay (WTP). The manufacturer also presented the cost-effectiveness acceptability frontier which suggests that glatiramer acetate and Rebif 22µg are the preferred options depending on the WTP.

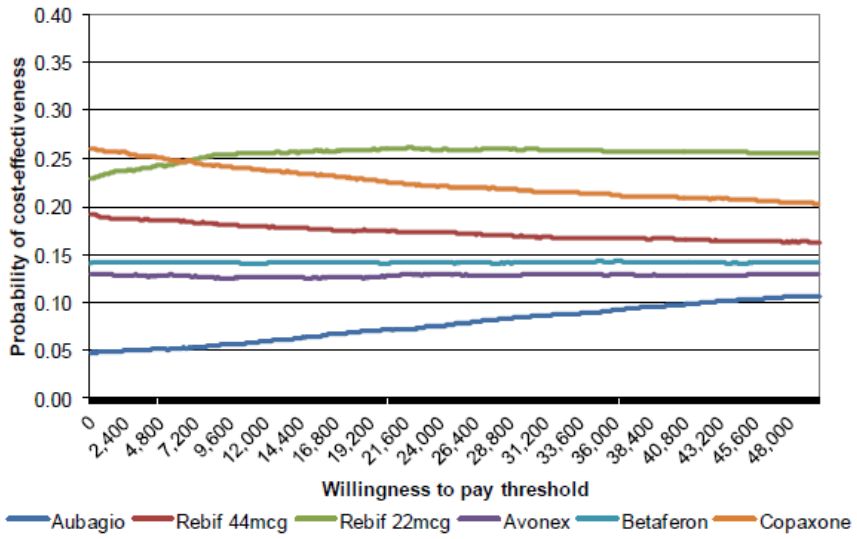


Figure 15: Multi-CEAC of teriflunomide versus all beta-interferons and glatiramer acetate (Manufacturer's submission, Figure B7.7.6)

The sensitivity analysis seemed to be well conducted. Instead of discarding transition matrices that predicted negative transition probabilities in the PSA, these negative predictions were made zero and the rest of the transitions adjusted. The uncertainty on treatment effects was included using the 95%

confidence intervals from the MTC, although there was no correlation modelled between the multiple treatment effects.

5.2.11.1 Patient access scheme

Patient access schemes (PAS) are arrangements which may be used on an exceptional basis for the acquisition of medicines for the NHS in England and Wales. These schemes help to improve the cost-effectiveness of a medicine and therefore allow NICE to recommend treatments which it would otherwise not have found cost-effective.

The manufacturer has submitted a price discount PAS. The proposed list price for teriflunomide 14mg per day is £13,529 per patient per year, however, the manufacturer is offering a fixed price to NHS patients in England and Wales of £[redacted] per patient, per year or £[redacted] per 28 tab pack, representing a discount of [redacted]% of the proposed list price.

This simple scheme will be valid for the whole licensed population. The scheme will apply from the point at which a Hospital Trust purchases a supply of teriflunomide for an NHS patient.

Using the PAS price in the cost-effectiveness model the manufacturer presents the base case probabilistic results versus the blended comparator (Table 48).

Table 48: Base case probabilistic cost-effectiveness results with and without the patient access scheme (corrected model)

	Without PAS		With PAS	
	Blended comparator	teriflunomide	Blended comparator	teriflunomide
Total costs (£)	[redacted]	[redacted]	[redacted]	[redacted]
Difference in total costs (£)		[redacted]		-£5,051
QALYs	[redacted]	[redacted]	[redacted]	[redacted]
QALY difference		[redacted]		0.1953
ICER (£)		[redacted]		Dominates

In the full incremental analysis using the corrected model and the PAS price teriflunomide was found to dominate all of the other comparators (i.e. less costly and more effective).

The manufacturer undertook multiple scenario and sensitivity analyses using the manufacturer's original uncorrected model. The results were highly sensitive to the hazard ratio on disability progression and were somewhat sensitive to discontinuation rates. In the uncorrected deterministic model the one-way sensitivity and structural analyses resulted in teriflunomide dominating the blended comparator in all scenarios except when the hazard ratio of teriflunomide is at its lower 95% confidence interval where the ICER compared to the blended comparator was £20,613 per QALY gained. The manufacturer also tested the inclusion of pre-2000 publications in the MTC, this resulted

in a deterministic ICER of £48,635 per QALY gained in the full incremental analysis using the PAS price compared to £[redacted] per QALY gained using the non-PAS price. The inclusion of the pre-2000 publications in the MTC resulted in a probabilistic ICER or £86,866 per QALY gained using the PAS price.

5.2.12 Subgroup analysis

Two subgroup analyses were undertaken by the manufacturer, the first, a comparison of teriflunomide versus fingolimod in HA patients, and the second, a comparison of teriflunomide versus natalizumab in RES patients.

The treatment effect of teriflunomide for comparison to fingolimod was estimated from TEMSO in patients who had at least one relapse within one year before randomisation, and with previous interferon use for at least 12 months, with a gap in DMT use of less than 12 months and at least one gadolinium-enhanced lesion at baseline. To compare to natalizumab, the treatment effect of teriflunomide was estimated from TEMSO in patients who had at least two relapses within one year of randomisation, and at least one gadolinium-enhanced lesion at baseline. Indirect comparisons were performed to determine the relative effectiveness of the treatments in the subgroups. Only 11 patients on teriflunomide 14mg had HA in TEMSO, but the relative risks of relapse and hazard ratio on progression were much lower than fingolimod, ([redacted]) and ([redacted]) respectively. The fingolimod effectiveness was based on the EPAR subgroup analysis considering patients who had at least one relapse within one year before randomisation, had a MS drug from at least 6-months, were still on treatment in the year before starting study drug. Only 33 patients on teriflunomide 14mg had RES in TEMSO and the relative risk of relapse was slightly lower ([redacted]) but the hazard ratio on progression was higher ([redacted]). The natalizumab effectiveness was based on the manufacturer's submission to NICE (TA 127) considering patients who had at least two relapses with one year before randomisation, and at least one GAD lesion at baseline.

The probabilistic cost-effectiveness analysis of teriflunomide versus fingolimod resulted in teriflunomide having lower costs and higher QALYs than fingolimod even when the price of fingolimod was lowered to £11,000 as an estimate of its PAS price.

The probabilistic cost-effectiveness analysis of teriflunomide versus natalizumab resulted in teriflunomide having lower costs and lower QALYs, but being the cost-effective treatment at a threshold of £30,000 per QALY gained.

The ERG does not consider the subgroup analyses reliable given the very small number of patients included in each of the teriflunomide groups and that the relative risks and hazard ratios calculated from the specified subgroups was combined with the natural history of the of the full RRMS

population. Furthermore only results from the TEMSO trial were used to calculate the teriflunomide effects.

5.2.12.1 Patient access scheme applied to subgroups

Using the PAS price of teriflunomide the manufacturer reports that teriflunomide dominates fingolimod in the HA population and teriflunomide is less costly and less effective than natalizumab with teriflunomide being the cost-effective treatment at a £30,000 per QALY gained threshold.

5.2.13 Model validation and face validity check

The manufacturer states that two independent health economists were involved in designing and building the model and ratifying the plausibility of results.

To explore the internal validity of the model the manufacturer compared results from the TEMSO trial with the model results (Table 49). The model predicted approximately twice the number of relapses in the model than in the TEMSO trial and the model predicted a -0.103 change in HRQoL compared to -0.006 change in HRQoL in TEMSO. The manufacturer noted the difference in the model and trial results and stated that differences were due to the variation in severity of the trial population and the modelled cohort. To investigate this further the ERG requested more validation in the points for clarification.

Table 49: Summary of teriflunomide base case model results compared with clinical data (Manufacturer's submission, Table B.7.7.1)

Outcome tested	Specific parameter	Clinical trial results	Model result
Disability/ Symptoms of multiple sclerosis/ Freedom of disease activity	Change in mean EDSS from baseline	TOWER 0.038 (1.4%) (at week 108) TEMSO 0.339 (4.9%) (at week 96)	0.53 (14.5%) (at year 2)
Relapse rate/ Severity of relapse / Resource utilisation hospitalisation	Average number of relapses leading to hospitalisation per patient in 2 years	TEMSO 0.111 (n=40 relapses)	0.207
	Average number of relapses not leading to hospitalisation per patient in 2 years	TEMSO 0.522 (n=187 relapses)	1.167
Health related quality-of-life	Change in EQ-5D from baseline	TEMSO -0.006 (-0.82%) (at week 96)	-0.103 (-17.2%) (at week 104)
Mortality	2-year mortality	0.0028% (No deaths occurred in TEMSO, 2 deaths in TOWER)	0.0039%

As part of the points for clarification the ERG asked the manufacturer to compare the results of the trial population to the modelled results of a population using the same characteristics as the trial population. In the teriflunomide arm the change in baseline EDSS scores and the change in EQ-5D scores were approximately [redacted] for the modelled results compared to the trial results (Table 50).

Table 50: Comparison of clinical trial and model results (Manufacturer's response, Table B10)

Parameter	Clinical trial results Pooled TEMSO, TOWER & TENER		Model results per patient	
Teriflunomide 14 mg	**Year 1**	**Year 2**	**Year 1**	**Year 2**
Average number of relapses	[redacted]	[redacted]	[redacted]	[redacted]
Mean EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Change in baseline EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Mean EQ-5D score	[redacted]	[redacted]	[redacted]	[redacted]
Change in baseline EQ-5D score	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	**Year 1**	**Year 2**	**Year 1**	**Year 2**
Average number of relapses	[redacted]	[redacted]	[redacted]	[redacted]
Mean EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Change in baseline EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Placebo	**Year 1**	**Year 2**	**Year 1**	**Year 2**
Average number of relapses	[redacted]	[redacted]	[redacted]	[redacted]
Mean EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Change in baseline EDSS score	[redacted]	[redacted]	[redacted]	[redacted]
Mean EQ-5D score	[redacted]	[redacted]	[redacted]	[redacted]
Change in baseline EQ-5D score	[redacted]	[redacted]	[redacted]	[redacted]

*Expected error from manufacturer

The model also predicts life-years approximately 6 times higher than the predicted QALYs. Part of this is because of the negative HRQoL in the more severe disability states (EDSS 8 and 9), however further exploration is needed to determine whether this large difference in modelled life-years and QALYs is particular to this model or has been seen in previous NICE appraisals.

Further validation is also needed to understand how the costs and QALYs of the treatments compare to previous cost-effectiveness analyses. This will be undertaken in the following section.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

From the manufacturer's validation and validity check (Section 5.2.13), the model does not correspond well with the results of the trial data, and the ERG are concerned that patients in the model transition too quickly through EDSS, transition too quickly to SPMS, have very low HRQoL and result in too few total QALYs. Because the ERG is concerned with the model validity a number of external validity checks have been undertaken. In this section we compare the results of from previous technology appraisals with the costs and QALYs produced by the manufacturer's model. The ICER is calculated for each treatment compared to BSC to make the ICERs comparable across appraisals.

It is evident from these results that the total QALYs accrued in the manufacturer's model are significantly lower than those QALYs reported in the technology appraisal (TA 32) for which the patient population is similar (RRMS patients non-HA/RES). The ERG underwent further exploration in an attempt to explain the differences and justify the model structure. The ERG's analysis focused

on three specific issues, HRQoL, the initial patient distribution of EDSS (i.e. severity) and natural history.

5.3.1 External validation

Three previous NICE appraisals of beta-interferons and glatiramer acetate (TA 32), natalizumab (TA 127) and fingolimod (TA 254) have estimated the costs and QALYs for patients with MS. Of particular relevance is TA 32 as this represents the same population and includes the same comparators used in the manufacturer's base case analysis of teriflunomide. Since teriflunomide was not included in previous appraisals all treatments have been compared to BSC to allow comparability. TA 32 found ICERs of beta-interferons and glatiramer acetate compared to BSC in the range of £48,130 - £98,171 per QALY gained in the RRMS population (Table 51). The QALYs in TA 32 range from 8.93 to 9.57 and the cost from £68,453 to £102,133. The results of TA 127 are similar to TA 32 although the estimated QALYs are slightly lower ranging from 6.01 to 7.58. The lower QALYs in TA 127 are not completely unexpected as this appraisal considers natalizumab for RES patients, a more severe population. TA 254 uses time horizons more than twice as long as the previous TAs and estimates higher costs of £224,311 to £321,730, however the range of QALYs is lower, 3.81 to 4.88. Lower QALYs in this appraisal may also be explained by the more severe population, HA patients. The current analysis is most similar to TA 254 having used the same natural history data (although not the exact same transition matrix) and time horizon; the current analysis has the highest range of costs and the lowest range of QALYs. The low QALY results are not expected as teriflunomide is expected to be used as first-line treatment. Further analysis was undertaken to understand why the QALYs estimated in this model are so low.

Table 51: Base case results of NICE appraisals for RRMS

	ID548 (current appraisal)*				TA 32			TA 127			TA 254		
	LY	QALY	Cost	ICER comparing to BSC	QALY	Cost	ICER comparing to BSC	QALY	Cost	ICER comparing to BSC	QALY	Cost	ICER comparing to BSC
BSC	[redacted]	[redacted]	£471,420	-	8.93	£68,453	-	6.15	£79,200	-	3.81	£224,311	-
IFN-β 1a (Avonex 30µg)	[redacted]	[redacted]	£493,409	£175,918	9.84	£112,251	£48,130	6.65	£119,200	£ 80,000	3.98	£271,646	£278,441
IFN-β 1b	[redacted]	[redacted]	£501,845	Dominated	9.95	£110,097	£40,827	-	-	-	-	-	-
IFN-β 1b	[redacted]	[redacted]	-	-	9.57	£102,133	£52,625	-	-	-	-	-	-
IFN-β 1a (Rebif 22g)	[redacted]	[redacted]	£489,703	£82,098	9.68	£112,796	£59,124	-	-	-	-	-	-
IFN-β 1a (Rebif 44µg)	[redacted]	[redacted]	£488,069	£79,310	9.73	£131,455	£78,752	-	-	-	4.13	£258,458	£106,709
Natalizumab	[redacted]	[redacted]	£528,364	£78,068	-	-	-	7.58	£159,500	£ 56,154	-	-	-
Glatiramer acetate	[redacted]	[redacted]	£486,555	£142,703	9.27	£101,831	£98,171	6.01	£110,000	Dominated	-	-	-
Fingolimod	[redacted]	[redacted]	£528,359	£133,254	-	-	-	-	-		4.88	£321,730	£ 91,046
Teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	-	-	-	-	-	-	-	-	-
Teriflunomide (PAS price)	[redacted]	[redacted]	[redacted]	£31,321									

* deterministic results as LY and BSC are not reported in the probabilistic analysis

5.3.2 Exploration of QALYs

5.3.2.1 Health-related quality-of-life

The differences in HRQoL estimates used in this and previous technology appraisals was explored and tested to determine the effect on the total QALYs accrued in the model. Table Y demonstrates that there is a wide range of HRQoL estimates available in the literature and applied in previous technology appraisals for EDSS health states. The difference is particularly large in the more severe EDSS states (i.e. 8 and 9), where HRQoL scores could be as low as -0.195 or as high as 0.232 in EDSS 9. The switch from a negative to positive value has a particularly large effect on the total QALYs. The manufacturer used the lowest HRQoL reported, as did TA 254, this may be the reason for the low QALYs. However, even when the highest HRQoL was used for each health state the overall additional QALYs gained was marginal. The QALYs for BSC increased by [redacted].

5.3.2.2 Initial patient distribution

The difference in initial patient distribution was also explored to determine its effect on the total QALYs gained. As discussed previously the initial distribution of EDSS patients used in the model was the same as the distribution of those patients participating in the risk-sharing scheme. This includes 4.5% of patients who are EDSS 7 and do not receive treatment in the model as per the ABN guidelines. In the model 50.5% of patients have an EDSS score of 4 or greater. The total QALYs in the BSC arm increases from [redacted] if patients are assumed to start treatment in EDSS 0. As expected the patients who start in lower disability states accrue more QALYs over the modelled time horizon. The initial EDSS distribution of patients differed across the NICE appraisals. TA 32 and TA 254 used more severe populations similar to the RSS population used in this model. TA 127 used distribution of the placebo arm of the AFFIRM trial. The difference in the initial EDSS distribution may explain some of the differences in the total QALYs between appraisals.

5.3.2.3 Natural history of progression

Patients who progress more quickly through the model will experience the negative HRQoL of being in the later disability states. This will also affect the total QALYs accrued by patients. As discussed previously the base case natural history data does not allow patients to improve their disability state, only stay the same or get worse. Natural history data estimated from the BSC arms of the TEMSO and TOWER trials allows patients to improve their disability state (i.e. patients can both progress and regress in terms of EDSS scores). The use of the trial natural history was used to test the effect on the total QALYs. The QALYs in the BSC arm increased from [redacted] allowing for both progression and regression in disability states. Unfortunately natural history matrices were not available for each of the previous NICE appraisals so it is unknown whether differences in the natural

history may be causing these differences between models. This is not expected to be the case as all of the models relied on the same data source, London Ontario data, although the estimated transition matrices differed due to difference in calculations or populations define, also some appraisals such as TA127 relied more on the trial data.

All previous appraisals used the London Ontario data for transitions from RRMS to SPMS and for transitions between EDSS within SPMS. As discussed later in Section 5.3.3, the calculation used by the manufacturer to calculate conversion to SPMS appears inconsistent with the natural history data available. However, as the same SPMS conversion data has been used in previous appraisals this in itself does not explain the low QALYs. Further, correcting the calculation and using an approach the ERG considered more in line with the natural history data, does not meaningfully change the total QALYs.

5.3.2.4 Conclusions of the exploration of QALYs

Multiple variables were tested to identify possible reasons for the low total QALYs estimates reported in the model. We tested the HRQoL, the initial patient distribution and the natural history of progression. The only scenarios that independently had a meaningful effect on the total QALYs were the use of the trial natural history data for the EDSS transitions and lowering the initial distribution of EDSS. This resulted in the BSC arm increasing to [redacted] QALYs, when using the trial data, and [redacted] QALYs, when initial EDSS was lowered to zero. The fact that using the trial natural history data and using a lower initial EDSS distribution resulted in total QALYs more similar to TA 32, support the ERG's preferred analysis that will be discussed in Section 6.

Table 52: HRQoL by EDSS Data from UK EQ-5D Studies

EDSS state	Health-related quality-of-life estimates							
	Parkin et al., 1998*	Orme et al., 2007*		Biogen Idec UK and Elan Pharma International, 2007		Evidence Review Group Report TA No. 127		TEMSO
		RRMS	SPMS	RRMS	SPMS	RRMS	SPMS	RRMS
0	—	0.870	0.825	0.91	0.87	0.959	0.874	[redacted]
1	—	0.799	0.754	0.84 (EDSS 0.5-1)	0.8 (EDSS 0.5-1)	0.688 (EDSS 1)	0.603 (EDSS 1)	[redacted]
2	—	0.705	0.660	0.74 (EDSS 1.5-2)	0.7 (EDSS 1.5-2)	0.688 (EDSS 1.5-2)	0.603 (EDSS 1.5-2)	[redacted]
3	0.71	0.574	0.529	0.61 (EDSS 2.5-3)	0.57 (EDSS 2.5-3)	0.645 (EDSS 2.5-3)	0.560 (EDSS 2.5-3)	[redacted]
4	0.66	0.610	0.565	0.65 (EDSS 3.5-4)	0.61 (EDSS 3.5-4)	0.610 (EDSS 3.5-4)	0.527 (EDSS 3.5-4)	[redacted]
5	0.52	0.518	0.473	0.56 (EDSS 4.5-5)	0.51 (EDSS 4.5-5)	0.581 (EDSS 4.5-5)	0.496 (EDSS 4.5-5)	[redacted]
6	0.49	0.460	0.415	0.49 (EDSS 5.5-6)	0.45 (EDSS 5.5-6)	0.538 (EDSS 5.5-6)	0.453 (EDSS 5.5-6)	[redacted]
7	0.35	0.297	0.252	0.44 (EDSS 6.5-7)	0.39 (EDSS 6.5-7)	0.477-0.343 (EDSS 6.5-7)	0.392-0.258 (EDSS 6.5-7)	-
8	—	–0.049	–0.094	–0.01 (EDSS 7.5-8)	–0.05 (EDSS 7.5-8)	0.343-0.232 (EDSS 7.5-8)	0.258-0.147 (EDSS 7.5-8)	-
9	—	–0.195	–0.240	–0.15 (EDSS 8.5-9.5)	–0.19 (EDSS 8.5-9.5)	0.232 to –0.135 (EDSS 8.5-9.5)	0.147 to –0.220 (EDSS 8.5-9.5)	-

5.3.3 Further exploratory analyses

Several additional univariate scenarios were explored by the ERG. The purpose of these scenarios is to illustrate the (independent) effect that selected model inputs have on the model outputs (costs and QALYs). This is to aid the Committee to understand how the model works, identify key drivers of cost-effectiveness and to help internally validate the model. It should be noted that the additional scenarios are exploratory in nature and hence do not necessarily reflect the ERG's view about the most plausible values of the model inputs. The ERG base case is developed in Section 6. Table 53 describes the scenarios undertaken by the ERG.

Table 53: A description of the exploratory analyses undertaken by the ERG

Label	Scenario	Description
Manufacturer deterministic corrected base case (blended comparator)	1	The base case deterministic analysis from the manufacturer's corrected model comparing to the blended comparator.
Basecase, corrected model (Rebif comparator)	2	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg.
Active LO data Natural history	3	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg using the 'active RRMS' population to calculate the EDSS transition matrix from the London Ontario data.
Trials natural history	4	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg using the placebo arm of the TEMSO and TOWER trial data to calculate the EDSS transition matrix.
No transition to SP allowed	5	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg with no transitions to SPMS in the model.
No transition to SP + trial natural history	6	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg using the placebo arm of the TEMSO and TOWER trial data to calculate the EDSS transition matrix and no transitions to SPMS in the model.
No relapse	7	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg with no relapses in the model.
Progression rate teriflunomide= Rebif 44µg	8	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg and the hazard ratio of teriflunomide is equal to the hazard ratio of Rebif 44µg in the MTC.
Progression rate Rebif 44µg = teriflunomide	9	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg and the hazard ratio of Rebif 44µg is equal to the hazard ratio teriflunomide of in the MTC.
Progression Rebif 44µg v teriflunomide from trial (TENERE HR = 0.98)	10	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg and the hazard ratio of teriflunomide is 0.98 (the direct trial result) compared to Rebif 44µg.
No waning of withdrawal after 2 years (default is 50% waning)	11	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg with no waning of withdrawal.
50% waning of treatment effect 2 years (default is no waning)	12	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg with 50% waning of treatment effects.
No withdrawal	13	The base case deterministic analysis from the manufacturer's

Label	Scenario	Description
		corrected model comparing to Rebif 44µg with no withdrawal.
Withdrawal rate of Rebif 44µg = teriflunomide	14	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg with withdrawal of Rebif 44µg equal to teriflunomide.
HRQOL from TEMSO	15	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg using the HRQoL data from the TEMSO trial for EDSS health states 0-7.
HRQOL of EDSS states 7, 8 and 9 are the same (currently less than zero)	16	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg eliminating the negative HRQoL states by making the HRQoL of EDSS states 8 and 9 equal to EDSS state 7.
Base EDSS 0 -6	17-23	The base case deterministic analysis from the manufacturer's corrected model comparing to Rebif 44µg and 7 scenarios starting all patients in the same EDSS score, from 0 to 6.

The manufacturer`s deterministic base case estimated an ICER of £[redacted] per QALY gained (scenario 1) (Table 54). The method used to construct the blended comparator may be biased and may obscure the workings of the model. In order to show the Committee more clearly how the model works, the ERG in all the additional analyses have compared teriflunomide with Rebif 44µg as it was used in the head-to-head trial of teriflunomide (scenario 2). The ICER of teriflunomide versus Rebif 44µg is £[redacted] per QALY gained.

The ERG conducted 6 analyses varying the natural history of progression. Scenario 3 uses the transitions calculated by the manufacturer for active RRMS using the London Ontario data. Scenario 4 uses the transitions calculated from the TOWER and TEMSO trials. Scenario 5 disallows transitions from RRMS to SPMS. Scenario 6 uses the trials natural history and disallows transitions from RRMS to SPMS. Scenario 6 increases the ICER for teriflunomide versus Rebif 44µg to £[redacted] per QALY. This is because under this scenario, patients move much more slowly through EDSS states and transitions from worse EDSS to better EDSS is permitted. To illustrate this, under the base case (scenario 2) 57% of patients are predicted to be in EDSS>6 after 10 years, while this percentage falls to 3% under scenario 6.

Scenario 7 shows the ICER is unaffected by the relapse rate, since the rate is low in absolute terms and similar between the treatments. Scenarios 8, 9 and 10 vary the treatment effect for progression of disease. Under scenario 8, the treatment effect for teriflunomide versus placebo is equal to the effect of Rebif 44µg versus placebo and for scenario 9, Rebif 44µg equals teriflunomide. Under scenario 10, the treatment effect for progression for teriflunomide versus Rebif 44µg is taken directly from the TENERE trial (hazard ratio 0.98), rather than from the MTC (hazard ratio 0.9).

Scenario 11 and 12 shows the effect of "waning" the treatment effect parameters for withdrawal and progression respectively. Scenario 12 shows that the ICER increases to £[redacted] per QALY gained if the treatment effects for teriflunomide versus placebo and Rebif 44µg versus placebo reduce by half

after 2 years. Scenarios 13 and 14 show the relative rate of withdrawals does not materially affect the ICER. Scenarios 15 and 16 show the effect of varying the HRQOL parameters related to EDSS states. Using the HRQOL parameters from the TEMSO trial reduces the ICER. However, it should be noted that the trial did not estimate HRQOL for severe EDSS states (greater than 6) hence the MS used the parameters from Orme (2007), which estimated a large decrease in HRQOL from state 6 to state 7 and negative HRQOL in states 8 and 9. Scenario 16 set the HRQOL for EDSS 8 and 9 equal to those for state 7, and the ICER increased to £[redacted] per QALY.

Scenarios 17 to 23 show the impact of changing the initial severity from EDSS =0 to EDSS = 6. Increasing the baseline severity reduces the difference in costs between the treatments and reduces the difference in QALY, because patients have less time before they reach disabling EDSS states and are withdrawn from drugs. However, there is no clear pattern on the ICER.

Each scenario was undertaken using the PAS price for most scenarios teriflunomide dominated Rebif 44µg. The only scenario for which teriflunomide did not dominate was scenario 6 that resulted in an ICER of £5,189 per QALY gained.

In conclusion, these sensitivity analyses have shown that the factors that appear to be the key drivers of cost-effectiveness are: 1) the choice of comparator, 2) the natural history and the rate of transition to SPMS, 3) the hazard ratio of progression and the "waning" of the treatment effect, 4) the HRQOL associated with the more severe health states, 5) and the possible "double counting" of withdrawal when estimating the relative effectiveness of the treatments. With the exception of the last factor, all these changes to the base case increase the ICER.

The purpose of this set of exploratory analyses was to understand the model by assessing whether the ICER moved in the expected direction to variable changes. For the most part this was the case. Contrary to general expectations the sensitivity analyses around withdrawal demonstrated that a treatment appears 'more' cost-effective when a treatment has higher withdrawal. In the base case patients that withdraw have BSC for all further lines of treatment. Since BSC is the cost-effective treatment in this analysis it is advantageous for treatment arms to get more patients on BSC more quickly. This provides the perverse result that increasing withdrawal rates lowers a treatments ICER versus a comparator. The fact that BSC is the most cost-effective treatment option calls into question the exclusion of BSC from the NICE scope.

Table 54: Additional sensitivity analyses by ERG

Scenarios		**Costs teriflunomide**	**Costs Rebif 44µg***	**QALY teriflunomide**	**QALY Rebif 44µg***	**Difference Costs**	**Difference QALY**	**ICER**	**ICER using PAS price**
Manufacturer deterministic corrected base case (blended comparator)	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base case, corrected model (Rebif comparator)	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Active LO data Natural history	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Trials natural history	4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
No transition to SP allowed	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
No transition to SP + trial natural history	6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£5,189
No relapse	7	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Progression rate teriflunomide= Rebif 44µg	8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Progression rate Rebif 44µg = teriflunomide	9	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Progression Rebif 44µg v teriflunomide from trial (TENERE HR = 0.98)	10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
No waning of withdrawal after 2 years (default is 50% waning)	11	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
50% waning of treatment effect 2 years (default is no waning)	12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
No withdrawal	13	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Withdrawal rate of Rebif 44µg = teriflunomide	14	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates

Scenarios		Costs teriflunomide	Costs Rebif 44µg*	QALY teriflunomide	QALY Rebif 44µg*	Difference Costs	Difference QALY	ICER	ICER using PAS price
HRQOL from TEMSO	15	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
HRQOL of EDSS states 7, 8 and 9 are the same (currently less than zero)	16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 0	17	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 1	18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 2	19	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 3	20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 4	21	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 5	22	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates
Base EDSS 6	23	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominates

* Except for scenario 1 for which the blended comparator is reported

5.4 Conclusions of the cost effectiveness section

In Section 5 the ERG have described the manufacturer's economic submission and discussed a number of issues which challenge the validity of the manufacturer's analysis. The manufacturer has positioned teriflunomide as a first-line treatment to be used for patients that would otherwise take beta-interferons or glatiramer acetates. They further state that fingolimod and natalizumab are not comparators of teriflunomide because these treatments have been approved to treat HA or RES patients. This suggests that teriflunomide will be used in non-HA/RES RRMS patients. The manufacturer has submitted no evidence for this subgroup, but the evidence submitted for the HA and RES subgroups suggest that teriflunomide is more effective in these populations meaning treatment may be less effective in the non-HA/RES population. As the treatment effect is unknown in this subgroup no specific analyses will be undertaken in section 6, however sensitivity of the model to the treatment effect will be considered generally. The ERG also found that the data used to populate the natural history and baseline characteristics used in the model represented a more severe population than is expected to be treated with first-line treatment. The ERG explores the use of trial data that the ERG considers to be more representative of patients to be initiated on first-line treatment.

The main comparator used in the model is a blended comparator. The ERG have commented that using a blended comparator does not meet the NICE reference case and the ERG use the fully incremental analysis in section 6 to form conclusions. The ERG continues to present results to the blended comparator only to facilitate comparisons with the manufacturer's submission.

The main source for the treatment effects used in this model is the manufacturer's post-2000 MTC. This MTC excluded trials pre-2000 due to the difference in diagnostic criteria used in MS clinical trials and the concern that pre-2000 trials may have included patients more severe than post-2000 patients. A separate MTC was undertaken including pre-2000 trials. The ERG discusses and explores the range of treatment effects from these MTCs and other sources in Section 6.

The ERG has discussed the manufacturer's inclusion of non-health costs in the model and the lack of detail provided on these costs. The ERG tests the exclusion of non-health costs in Section 6.

6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

6.1 Overview

This section details the ERG's further exploration of the issues and uncertainties raised in the review and critique of the manufacturer's cost-effectiveness analysis presented in Section 5. This section will focus on,

1. Population
2. Comparators
3. Natural history
4. Treatment effect
5. Health-related quality-of-life
6. Costs

In each further section the ERG's analyses compare the results of teriflunomide to Rebif 44µg and the blended comparator in order to make comparisons with the previous analyses presented by the ERG and manufacturer. All univariate analyses are undertaken using the manufacturer's corrected model using deterministic analyses. Deterministic analyses were used to show the effect of independent variable changes without the variation caused by the probabilistic analysis and because running the probabilistic analysis was more time consuming. Each scenario is also run using the PAS price for teriflunomide. In the final section the ERG selects its preferred univariate analysis in each previous section and combines them to present the ERG's preferred analysis. Both deterministic and probabilistic results are presented for the ERG's preferred analysis.

The univariate analyses undertaken by the ERG demonstrate the sensitivity of the model to the natural history of progression and the relative treatment effect of progression. The ERG's additional analyses found no scenarios in which teriflunomide was cost-effective at a £30,000 per QALY using the non-PAS price, however, using the PAS price teriflunomide was less costly and more effective in almost all univariate scenarios.

Multivariate analyses were undertaken by the ERG choosing a scenario the ERG considered to be more reflective of the patient population likely to receive teriflunomide and more externally valid. Using the non-PAS price teriflunomide had an ICER greater than £[redacted] per QALY or was more [redacted]. Using the PAS price teriflunomide was less costly and more effective than the blended comparator, but had an ICER greater than £86,000 per QALY gained compared to glatiramer acetate in the fully incremental analysis.

6.2 Additional ERG analyses

6.2.1 Population

The ERG expressed concerns about how well the population being modelled represents those who will be treated in clinical practice (Section 5.2.2). The modelled population represents all RRMS patients, although the population of interest to the manufacturer is those patients that would otherwise use beta-interferons or glatiramer acetate. Much of the data for the model was taken from the TEMSO trial which had a high proportion of HA and RES patients in the BSC and teriflunomide 14 mg arms (409/761). No evidence was provided for the population of RRMS patients without HA or

RES. It is not clear from the manufacturer's submission whether the population of interest (i.e RRMS non- HA/ RES) is well represented by the natural history data or treatment effects.

The RSS although representative of the current prevalent population may not represent well the population that are likely to be initiated on teriflunomide. The model used the initial distribution of EDSS from the RSS, this population had more severe initial disability than the trial populations. As teriflunomide is expected to be used in first-line, the ERG considered a less severe initial disability more reflective of the expected population who would initiate teriflunomide and chose to implement the initial EDSS distribution of the TEMSO and TOWER trials rather than the RSS.

The ERG have previously shown that the lower the initial EDSS the lower the ICER comparing teriflunomide to Rebif 44µg (Table 54). The manufacturer's model allowed for the initial distribution to be set to the initial distribution of the TEMSO and TOWER trials. The trial distribution had a higher percentage of patients in lower EDSS states and therefore lowered the ICER comparing teriflunomide to the blended comparator or Rebif 44µg (Table 55).

Table 55: The results comparing differences in the initial distribution of EDSS

Technology	Manufacturer's corrected base case			Initial distribution TEMSO + TOWER		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

6.2.2 Comparator

In Section 5.2.3 the ERG discusses the manufacturer's use of a 'blended' comparator. As discussed this blended comparator is calculated from the weighted average costs and treatment effects of the beta-interferons and glatiramer acetate. However, this method of calculation does not take into account the variation in consequences across the treatments. The ERG has corrected the calculation of the blended comparator to demonstrate the effect on the ICER. The ERG's calculation was done using the deterministic results from the corrected model. The ERG calculated a blended comparator by weighting the QALY and cost results from each treatment to take into account the potential variation in consequences that could be truncated by the manufacturer's calculation of the blended comparator (Table 56). The ERG's calculation of the blended comparator results in a deterministic ICER of £[redacted] per QALY gained (Table 57), compared to the manufacturer's calculation of the blended compared which results in a deterministic ICER of £[redacted] per QALY gained.

Table 56: Calculation of the blended comparator from the manufacturer's deterministic results

Technologies	Total costs (£)	Total QALYs	Market share
Glatiramer acetate	[redacted]	[redacted]	[redacted]
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	[redacted]
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	[redacted]
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	[redacted]
IFNβ-1b	[redacted]	[redacted]	[redacted]
Blended comparator*	[redacted]	[redacted]	[redacted]

*Calculated as the market share weighted average from the corrected model PSA results

Table 57: Deterministic results using the ERG's calculation of the blended comparator from Table 56

Technology	Manufacturer's corrected base case			ERG's calculation of the blended comparator		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Having shown what the ERG considers to be the correct method for calculating the blended comparator it is important to reassert that the ERG do not consider the blended comparator the correct comparator as it does not meet the NICE methods reference case of comparison to the next best alternative. Using the blended comparator also masks the effects of changes in the model as different treatments included in the blended comparator react differently and sometimes in opposite directions to changes. This may explain why sensitivity analysis around the blended comparator does not appear to change the ICER. The ERG's analysis will focus on the full incremental analysis but will also provide results for the blended comparator for comparison to the manufacturer's results.

The remaining analyses use the manufacturer's calculation of the blended comparator to maintain comparability with the manufacturer's submission.

6.2.3 Natural history

In the natural history section the ERG will discuss issues and possible corrections for disability progression and conversion to SPMS. In Section 5 the ERG also discussed relapse rates, but as there were no significant issues with relapse rates the ERG accepts the manufacturer's base case rates.

6.2.3.1 Disability progression

The base case EDSS transition probabilities used to model the patients disability progression were considered by the ERG to be from a population more severe than the first-line patient population in which teriflunomide is expected to be used. The data was also estimated without allowing for patients to improve their disability. Past technology appraisals (TA 254) have criticised use of the London

Ontario data because of the lack of transitions to better EDSS states. Further, the manufacturer demonstrates that this assumption is inconsistent with the trial data (Table 29). The manufacturer 's model allows a scenario to be tested where RRMS patients transition through EDSS using data from the placebo arms of the TEMSO and TOWER trials. This scenario is likely to be more representative of first-line patients and allows patients to improve their disability. The ERG demonstrates the effect of this model change (Table 58). The ERG's preferred analysis uses this trial data. When the PAS price for teriflunomide is used in this scenario the total costs of teriflunomide are £[redacted], resulting in teriflunomide being less costly and more effective than the blended comparator or Rebif 44µg.

Table 58: Results comparing disability progression in RRMS from London Ontario data (base case) to the BSC arms of the trials

Technology	Manufacturer's corrected base case			TEMSO + TOWER trial data disability progression for RRMS		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

*ICER of teriflunomide compared to selected technology

Using the disability progression from the trials in the model is an example of the blended comparator masking the expected effect on the ICER. The expected direction of effect is for the slower rate of transitions to increase the ICER as is the case in the comparison with Rebif 44µg, however when comparing to the blended comparator the ICER goes down (Table 58). This unexpected change may occur because one of the treatments used in the blended comparator is less effective than BSC.

The manufacturer does not provide any additional data for SPMS transitions through EDSS. The ERG test the effect of slowing transition through EDSS during SPMS and allowing patients to improve their disability scores by using the same trial transitions in SPMS that have been used in RRMS (Table 59). The ERG recognises that these transitions are likely to under-estimate disability transition of SPMS patients; however this analysis does demonstrate the sensitivity of the ICER to this variable and the direction of effect of assuming slower progression in SPMS.

Table 59: Results comparing disability progression in SPMS from London Ontario data to the BSC arms of the trials

Technology	Manufacturer's corrected base case			TEMSO + TOWER trial data disability progression for SPMS		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

*ICER of teriflunomide compared to selected technology

The results of Table 59 demonstrate that if patients progress through EDSS at the same rate whether they are RRMS or SPMS than the ICER of teriflunomide is higher. These results suggest that if patients with SPMS progress more slowly than is predicted in the London Ontario data than the manufacturer's base case ICER may be underestimated. There may be a case for this assumption given that the London Ontario data does not allow for improvement in EDSS and the progression rates for RRMS patients are lower from the trial than from the London Ontario data.

When the PAS price for teriflunomide is used in this scenario the total costs of teriflunomide are £[redacted], resulting in teriflunomide being less costly and more effective than the blended comparator or Rebif 44µg.

6.2.3.2 Conversion to SPMS

As discussed in Section 5.2.6.2, the manufacturer's estimates of conversion to SPMS do not predict 15 year conversion similar to those from external sources. Previous data suggests that the median time for conversion from EDSS 1 to SPMS is 15 years. From this, the manufacturer correctly estimates that the rate of transition is [redacted]). However, this is the average rate of conversion, and implicitly a proportion of patients will have passed through other EDSS states before converting to SPMS. The manufacturer incorrectly interprets this parameter as the rate of conversion from stage 1 EDSS to SPMS for a patient who does not progress through other EDSS states. Consequently, in the model, the rate of conversion increases exponentially as EDSS progresses. The effect in the model is to over predict the speed of conversion for a patient starting at EDSS 1. This can be seen by setting the initial distribution of all patients to EDSS 1 in the model. With the other natural history parameters set to London Ontario, the model predicts 52% will have converted to SPMS (under BSC) by 11 years, that is, the median time to conversion from EDSS 1 to SPMS is between 10 and 11 years, considerably less than 15 years.

The ERG have used an approximate method to estimate more appropriate values for the "intercept" rate of conversion to SPMS, which we define as the annual rate of conversion from stage 1 EDSS to SPMS for a patient who does not progress to other EDSS states. These values were estimated by

setting the initial EDSS to state 1, and then using the Excel goal-seek function to estimate the rate of transition to SPMS that would make 50% of patients convert to SPMS at 15 years under BSC treatment. The relative risks of converting to SPMS at other EDSS states, compared to EDSS 1 were kept at the base case values. Given the London Ontario natural history, the "intercept" is estimated to be 0.0263. Given the TEMSO / TOWER natural history, the intercept is estimated to be 0.0299 per year. With these intercept rates of conversion, 50% of patients who started in EDSS 1 would convert to SPMS by 15 years under best supportive care, that is the model is consistent with the natural history. Table 60 compares the manufacturer's estimates of transitions to SPMS with those calculated by the ERG. These conversion probabilities result in patients transitioning in and out of SPMS (from RRMS and to death) as pictured in Figure 17. The manufacturer's model estimates patients transition to SPMS more quickly than in the ERG's corrected scenarios.

Table 60: Estimates of conversion to SPMS by EDSS state

EDSS	Manufacturer's calculation of the probability of conversion to SPMS during following year	ERG calculation of probability of the conversion to SPMS during following year using natural history from trials	ERG calculation of probability of the conversion to SPMS during following year using natural history from London Ontario data
0	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]
6	[redacted]	[redacted]	[redacted]
7	[redacted]	[redacted]	[redacted]
8	[redacted]	[redacted]	[redacted]
9	[redacted]	[redacted]	[redacted]

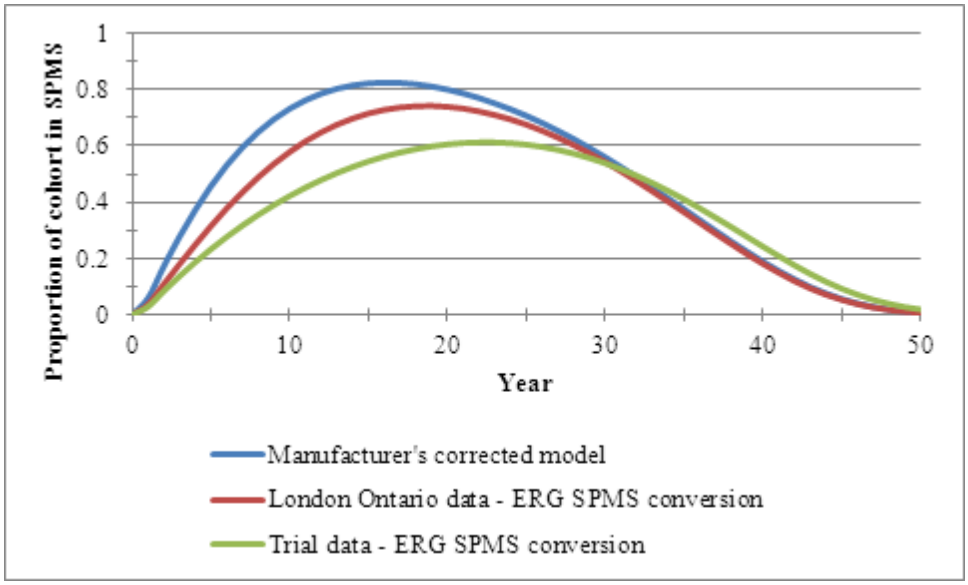


Figure 16: The proportion of patients in SPMS (ERG analysis based on the manufacturer model)

The ERG's calculations of the conversion probabilities do not result in the median population converting at 15 years since the base case model uses an initial distribution more severe than EDSS 1. It is only EDSS 1 patients who are expected to transition to SPMS at 15 years while patients in higher EDSS will transition more quickly. The difference in the proportion of patients in SPMS between the ERG's SPMS conversions using the London Ontario data for the natural history compared to using the trial data is not caused by the small difference in the conversion probabilities, but is caused by the difference in progression through EDSS (no regression in the London Ontario data versus regression in the trial data) which also affects the conversion to SPMS.

Changing the model to use the conversion to SPMS as calculated by the ERG resulted in only a small increase in the ICERs (Table 61). This scenario using the PAS price results in teriflunomide dominating (i.e. lower costs and higher QALYs) the blended comparator or Rebif 44µg.

Table 61: Results comparing conversion to SPMS as calculated by the manufacturer to those calculated by the ERG both using the London Ontario data

Technology	Manufacturer's corrected base case			ERG correct SPMS conversion using London Ontario natural history of progression		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

6.2.4 Treatment effect

The manufacturer's sensitivity analysis demonstrates that the ICER is most sensitive to the treatment effect on progression. The ERG has previously discussed issues relating to the manufacturer's estimated treatment effect (Section 5.2.7.1).

The results of the MTC have two surprising findings, firstly the MTC suggests that IFNβ-1b is less effective than placebo and secondly that the effect of teriflunomide compared to Rebif 44µg is better than the trial result. The ERG tests three scenarios. Firstly the ERG tests the assumption that the treatment effect of teriflunomide is [redacted] compared to Rebif 44µg (i.e. the treatment effect of teriflunomide versus placebo is 0.77) based on the direct head-to-head evidence from the TENERE trial. Secondly the ERG tests the assumption that there is no difference in effect between teriflunomide and Rebif 44µg (i.e. the treatment effect of teriflunomide versus placebo is 0.79). Finally the ERG test uses the relative effects calculated from the 'all studies' MTC. The resulting ICERs are higher than in the base case (Table 62). Despite less favourable treatment effects on relapse and equal progression effects compared to Rebif 44µg the total QALYs for teriflunomide are still higher than for Rebif 44µg due to the lower withdrawal rates. When the PAS price for teriflunomide is used the total costs of teriflunomide range from £[redacted] to £[redacted]; in all scenarios teriflunomide is less costly and more effective than the blended comparator or Rebif 44 µg.

Table 62: Results of comparing the relative treatment effect on progression of teriflunomide

Technology		Blended comparator	Rebif 44µg	Teriflunomide
Manufacturer's corrected base case (teriflunomide HR= 0.71)	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
TENERE trial effectiveness (teriflunomide HR= 0.77)	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
Same effectiveness as Rebif 44µg (teriflunomide HR= 0.79)	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
All Studies MTC (teriflunomide HR=0.69)	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

In Section 5, the ERG highlighted that despite other evaluations having used shorter time horizons, the use of a 50 year model time horizon in order to fully represent lifetime costs and effects associated with treatments is considered appropriate by the ERG. However, the manufacturer's base case assumes that those patients receiving DMTs have a constant and continued treatment effect, so long as they remain on treatment, over the 50 year time horizon of the model. The ERG considers this extrapolation assumption for the treatment effect - informed by 2 year trials - to be optimistic. In the manufacturer's sensitivity analysis, scenarios exploring possible waning of treatment effect were considered. They found that a 25% decrease in efficacy after 5 years increased the ICER compared to the blended comparator by £[redacted] per QALY, and a 50% decrease increased it by £[redacted] per QALY. The ERG for TA 254 included waning of the treatment effects in their preferred analysis to address uncertainty concerning the extrapolation period given that the time horizon was markedly longer than in previous MS appraisals (i.e. 50 versus 30 years).

6.2.5 Costs

The ERG remarked that the manufacturer's stated perspective of the analysis is the NHS and PSS perspective; however, additional non-health costs have been included in the model. The NICE reference case states that non- NHS and non-personal and social services costs should not be included in the calculation of the ICER.[32] To align the current analysis with the NICE Methods Guide the non-health costs were excluded from the analysis. Table 63 compares the difference in the results of excluding these costs, the ICER increases ~£[redacted] per QALY gained. Using the PAS price for teriflunomide the total costs of teriflunomide decrease to £[redacted] and teriflunomide is less costly and more effective when non-health costs are excluded.

Table 63: Results comparing the exclusion of non-health costs

Technology	Manufacturer's corrected base case			Excluding non-health costs		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

The ERG's preferred analysis excludes the non-health costs.

6.2.6 Health-related quality of life

The manufacturer's review of the HRQoL literature for MS excludes the previous technology appraisals. The ERG compares the HRQoL used in the model, Orme et al. 2007, to the HRQoL used in previous technology appraisals (Table 52). In Section 5.3.3 the ERG determined that the ICER is sensitive to the chosen HRQoL. The manufacturer's model uses the Orme et al. HRQoL as the base case. This data provides the lowest HRQoL values in the literature. The manufacturer's model also tests as a sensitivity analysis the use of the TEMSO data in the base case with the Orme et al. data for EDSS states 7-9. This scenario has higher HRQoL values in EDSS states 0-6, but still has negative values for 8 and 9. The ERG tested some additional scenarios,

1. TEMSO otherwise TA 127
2. TEMSO otherwise the average of all data from Table 52
3. The average of all data from Table 52
4. TEMSO otherwise the change between states from Orme et al. applied to TEMSO

Scenario 1 used the TEMSO data for EDSS states 0-6 in RRMS, since no HRQoL data was available from TEMSO for EDSS states 7-9 or SPMS, HRQoL from TA 127 was used for EDSS states 7-9 in RRMS and 0-9 in SPMS. In scenario 2 the TEMSO data was used for EDSS states 0-6 in RRMS and the other states were calculated as the average for each state from Table 52. Scenario 3 used the average for each state as calculated from Table 52. Scenario 4 used the TEMSO data for EDSS states 0-6 in RRMS and applied the difference between states 6, 7, 8 and 9 in the Orme data to calculate states 7, 8 and 9 from state 6 in the TEMSO data. The HRQoL for SPMS in scenario 4 was calculated by applying the difference between RRMS and SPMS for each state in the Orme et al. data and applying the differences to the RRMS already calculated.

Table 64: HRQoL by EDSS Data from UK EQ-5D Studies

EDSS state	Health-related quality-of-life estimates							
	1. TEMSO + TA 127		2. TEMSO + averages		3. Averages		4. TEMSO + differences from Orme et al.	
	RRMS	SPMS	RRMS	SPMS	RRMS	SPMS	RRMS	SPMS
0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
9	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The results of applying each of the 4 scenarios to the manufacturer's base case analysis are presented in Table 64. The ICERs of teriflunomide were highest when the average HRQoL for each health state was used. All scenarios increased the ICERs from the base case values. When the PAS price of teriflunomide is used the total costs of teriflunomide is £[redacted], this makes teriflunomide less costly and more effective than the blended comparator and Rebif 44µg in all of the HRQoL scenarios undertaken.

Table 65: Results of the ERG's HRQoL scenarios

Technology		Blended comparator	Rebif 44µg	teriflunomide
Manufacturer's corrected base case	QALYs	[redacted]	[redacted]	[redacted]
	Cost	[redacted]	[redacted]	[redacted]
	ICER*	[redacted]	[redacted]	[redacted]
1. TEMSO + TA 127	QALYs	[redacted]	[redacted]	[redacted]
	ICER*	[redacted]	[redacted]	[redacted]
2. TEMSO + averages	QALYs	[redacted]	[redacted]	[redacted]
	ICER*	[redacted]	[redacted]	[redacted]
3. Averages	QALYs	[redacted]	[redacted]	[redacted]
	ICER*	[redacted]	[redacted]	[redacted]
4. TEMSO + Orme differences	QALYs	[redacted]	[redacted]	[redacted]
	ICER*	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

The ERG selected to use scenario 4 in the ERG's preferred analysis. The ERG considered scenario 4 most representative of the patients being treated with teriflunomide as it uses the data from the TEMSO trial as the baseline estimates of HRQoL and estimates differences by EDSS state from a large UK based survey. However, the Orme study was of an older population (mean age of 51 versus a mean age of 39 from the TEMSO trial) and might not represent first-line RRMS patients. The Orme study also controlled for time since diagnosis and found a statistically significant increase in utility for each year of diagnosis. Time since diagnosis was not included in the manufacturer's model, which will bias the estimates of patients in later stages of disease. For these reasons scenario 4 was preferred to the manufacturer's base case HRQoL which came from the Orme study.

6.2.7 Sub-group analysis

The manufacturer reports that teriflunomide dominates fingolimod in the HA subgroup and is less costly and less effective, but cost-effective, compared to natalizumab in the RES subgroup. In Section 5.2.12 the ERG discussed the uncertainty of these results given the small patient numbers used to calculate the relapse rate and progression of teriflunomide in these subgroups, 11 patients in HA and 33 patients in RES.

The ERG presents comparisons of teriflunomide to natalizumab and fingolimod in the HA and RES populations assuming patients on teriflunomide have the same effectiveness as in the full RRMS population (Table 66). These results suggest that natalizumab and fingolimod are not cost-effective at a £30,000 per QALY threshold compared to teriflunomide in the RES and HA populations, respectively, when teriflunomide is no more effective in these populations than in the complete RRMS population. When the PAS price of teriflunomide is used then the total cost of teriflunomide

is £[redacted] and the ICERs for natalizumab and fingolimod are higher and not cost-effective compared to teriflunomide at a £30,000 per QALY threshold (Table 67).

Table 66: Deterministic analysis of the manufacturer's corrected base case using the RRMS results of teriflunomide and the subgroup results of natalizumab and fingolimod

Technology	QALYs	Cost	ICER	Population
teriflunomide	[redacted]	[redacted]	[redacted]	
natalizumab	[redacted]	[redacted]	[redacted]	RES
fingolimod	[redacted]	[redacted]	[redacted]	HA

Table 67: Deterministic analysis of the manufacturer's corrected base case using the RRMS results of teriflunomide and the subgroup results of natalizumab and fingolimod using the PAS price

Technology	QALYs	Cost	ICER	Population
teriflunomide	[redacted]	[redacted]	[redacted]	
natalizumab	[redacted]	[redacted]	[redacted]	RES
fingolimod	[redacted]	[redacted]	[redacted]	HA

6.2.8 ERG's preferred analysis

The ERG's preferred analysis contains the following changes to the manufacturer's base case,

1. Trial distribution of initial EDSS – to be more reflective of first-line patients
2. Trial estimates of natural history – includes improvements in progression and more reflective of first-line patients
3. ERG calculation of SPMS conversion – better external validity
4. Treatment effects from the 'All Studies MTC' – better external validity
5. Trial based HRQoL and changes from Orme et al. – more reflective of treated population
6. Excludes non-health costs – meets the NICE reference case

The ERG's preferred analysis results in QALYs much higher and costs much lower than those reported in the manufacturer's corrected base case analysis (Table 68). These results may have external validity as the results of the ERG's preferred analysis are slightly higher than the costs and QALYs reported in TA 32. This is expected as TA 32 used natural history data that did not allow for patients to improve their disability progression.

Table 68: Probabilistic results of the full incremental analysis comparing the ERG's preferred analysis, non-PAS price

Technology	Manufacturer's corrected base case			ERG's preferred analysis		
	QALYs	Cost	ICER	QALYs	Cost*	ICER
Glatiramer acetate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1b	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

* Not ordered by cost for easier comparison to the base case results

The resulting ICERs of the ERG's preferred analysis are similar to those from the manufacturer's corrected base case (Table 68). However, the QALYs are much higher and the costs much lower. The increase in QALYs is due to the ERG's assumption that patients are less progressed when they initiate on treatment (Section 6.2.1), that they may improve their EDSS (i.e. disability) (Section 6.2.3.1) and that HRQoL is more similar to the trial results (Section 6.2.6). The decrease in costs is mainly due to the exclusion of non-health costs (Section 6.2.5).

For comparison to previous analyses the ERG's preferred analysis was also used to compare teriflunomide to the blended comparator and Rebif 44µg (Table 69). The ICERs in the ERG's preferred analysis are higher than in the manufacturer's corrected base case, but the conclusions are the same at a £30,000 per QALY threshold.

Table 69: Probabilistic results comparing the ERG's preferred analysis using selected comparators, non-PAS price

Technology	Manufacturer's corrected base case			ERG's preferred analysis		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

The ERG's preferred analysis was also run using the manufacturer's PAS price of £[redacted]% of the proposed list price. The results of this analysis are reported in Table 70. These results suggest

that teriflunomide is more effective and more costly than glatiramer acetate with an ICER of £107,148 teriflunomide would not be considered cost-effective at a threshold of £30,000 per QALY. These results also suggest that Rebif 22µg is more effective and more costly than teriflunomide, but not cost-effective with an ICER of £917,667. However, teriflunomide does dominate the blended comparator and Rebif 44µg (Table 71).

Table 70: Probabilistic results of the full incremental analysis comparing the ERG's preferred analysis with the PAS price

Technology	Manufacturer's corrected base case with PAS price			ERG's preferred analysis with PAS price		
	QALYs	Cost	ICER	QALYs	Cost*	ICER
Glatiramer acetate	[redacted]	[redacted]	-	[redacted]	[redacted]	-
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	£917,667
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1b	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
teriflunomide	[redacted]	[redacted]	Dominates	[redacted]	[redacted]	£107,148

* Not ordered by cost for easier comparison to the base case results

Table 71: Probabilistic results comparing the ERG's preferred analysis using selected comparators, with the PAS price

Technology	Manufacturer's corrected base case with PAS price			ERG's preferred analysis with PAS price		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	Teriflunomide Dominates	[redacted]	[redacted]	Teriflunomide Dominates
Rebif 44µg	[redacted]	[redacted]	Teriflunomide Dominates	[redacted]	[redacted]	Teriflunomide Dominates
teriflunomide	[redacted]	[redacted]		[redacted]	[redacted]	-

*ICER of teriflunomide compared to selected technology, **Are the same by chance

Deterministic results of the ERG's preferred analysis are presented in Table 72 to Table 75. The deterministic results, although different, support the conclusions drawn from the probabilistic results.

Table 72: Deterministic results comparing the ERG's preferred analysis with the non-PAS price

Technology	Manufacturer's corrected base case			ERG's preferred analysis with the non-PAS price		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Glatiramer acetate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFNβ-1b	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 73: Deterministic results comparing the ERG's preferred analysis using selected comparators, with the non-PAS price

Technology	Manufacturer's corrected base case			ERG's preferred analysis		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Rebif 44µg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
teriflunomide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ICER of teriflunomide compared to selected technology

Table 74: Deterministic results comparing the ERG's preferred analysis with the PAS price

Technology	Manufacturer's corrected base case with the PAS price			ERG's preferred analysis with the non-PAS price		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Glatiramer acetate	[redacted]	[redacted]	-	[redacted]	[redacted]	-
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1b	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
teriflunomide	[redacted]	[redacted]	Dominates	[redacted]	[redacted]	£86,946

Table 75: Deterministic results comparing the ERG's preferred analysis using selected comparators, with the PAS price

Technology	Manufacturer's corrected base case with PAS price			ERG's preferred analysis with PAS price		
	QALYs	Cost	ICER*	QALYs	Cost	ICER*
Blended comparator	[redacted]	[redacted]	Teriflunomide Dominates	[redacted]	[redacted]	Teriflunomide Dominates
Rebif 44µg	[redacted]	[redacted]	Teriflunomide Dominates	[redacted]	[redacted]	£3,318
teriflunomide	[redacted]	[redacted]	-	[redacted]	[redacted]	-

*ICER of teriflunomide compared to selected technology

Given the remaining uncertainty about which MTC is most appropriate the ERG include below the results of the ERG's preferred analysis excluding the 'all studies' MTC and using the post-2000 MTC (Table 76). The ERG found that using the ERG's preferred analysis except the 'all studies' MTC resulted in teriflunomide being cost-effective at a threshold of £30,000 per QALY gained.

Table 76: Deterministic results comparing the ERG's preferred analysis with the PAS price but using the 'post-2000' MTC.

Technology	ERG's preferred analysis with the PAS price and all studies MTC			ERG's preferred analysis with the PAS price and post-2000 MTC		
	QALYs	Cost	ICER	QALYs	Cost	ICER
Glatiramer acetate	[redacted]	[redacted]	-	[redacted]	[redacted]	-
IFNβ-1a 44µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1a 22µg (Rebif)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1a (Avonex 30µg)	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
IFNβ-1b	[redacted]	[redacted]	Dominated	[redacted]	[redacted]	Dominated
teriflunomide	[redacted]	[redacted]	£86,946	[redacted]	[redacted]	£6,266

6.3 Conclusions from ERG analyses

In Section 6 the ERG has explored a number of additional analyses to highlight the uncertainty around the manufacturer's estimated ICER and to demonstrate additional model results. In Section 6.2.8 the ERG has undertaken univariate analyses and presented the results of the ERG's preferred analysis. This analysis combined population characteristics, HRQoL values and natural history data from the TEMSO trial that the ERG considers to be more reflective of the first-line patients eligible for teriflunomide. The ERG recalculated the SPMS conversion and excluded non-health costs. Finally the ERG chose to use the MTC results which included all trials, pre and post-2000.

The multiple univariate analyses undertaken demonstrate the robustness of the conclusions, that when using the non-PAS price teriflunomide [redacted]. When using the PAS price teriflunomide was generally found to be cost-effective at £30,000 per QALY, except the ERG's preferred analysis.

Despite the significant differences in the total QALYs and costs estimated in the manufacturer's base case and the ERG's preferred analysis the overall cost-effectiveness conclusions are relatively similar. In the manufacturer's base case and the ERG's preferred analysis teriflunomide was dominated by Rebif 22µg (i.e. more costly and less effective) using the non-PAS price; comparing to Rebif 44µg or the blended comparator, teriflunomide was [redacted] at £30,000 per QALY gained. Using the PAS price the manufacturer's analyses result in teriflunomide dominating its comparators, however using the ERG's preferred analysis and the PAS price teriflunomide is not cost-effective at £30,000 per QALY gained. This conclusion depends on whether the pre-2000 trials are included in the MTC.

Further exploration of the ERG's preferred scenario found that teriflunomide was cost-effective when the post-2000 MTC results were combined with the other ERG preferences (Table 76). There is still a high degree of uncertainty around which MTC analysis is the most appropriate. The pre-2000 trials were excluded by the manufacturer because of the different selection criteria used to define MS. The manufacturer was concerned that patients included in the pre-2000 trials would be more severe than those in the post-2000 trial. However, even when only the post-2000 trials are included a large amount of heterogeneity remains given the large difference in the proportion of HA and RES patients within the trials. The ERG chose the 'All studies' MTC because the trial results seemed to be more face valid given that the hazard ratio between teriflunomide and Rebif 44µg was more similar to the head-to-head trial and the hazard ratio on Betaferon compared to placebo was not greater than one.

Additional uncertainty remains around the transition of patients in SPMS. The ERG demonstrated that when SPMS patients progress more slowly through EDSS the ICER for teriflunomide increases. However this could not be incorporated in the ERG's preferred analysis due to the lack of alternative choices in transition matrices. Hence there may be remaining bias in the model which favours teriflunomide.

Treatment effect may also be biased in favour of teriflunomide by including HA and RES patients in the estimates of effectiveness. The ERG could not account for this bias due to a lack of data on the patient population that excludes HA and RES patients.

7 End of life

End of life considerations are not relevant to the appraisal of teriflunomide for RRMS.

8 Overall conclusions

Most importantly to the ICER and the overall conclusions is the remaining uncertainty regarding the size of the comparative effect of teriflunomide and the most appropriate methods for the MTC. Teriflunomide has demonstrated a small but statistically insignificant effect on 3 month SAD compared to Rebif 44µg. However the 6 month SAD results do not demonstrate a statistically significant treatment effect from placebo. The economic analysis relies on a MTC of 3 month SAD results.

The ERG has demonstrated that the comparative treatment effect on progression is the most important input in the cost-effectiveness model. The manufacturer uses a MTC of trials post-2000 to estimate treatment effects on progression. The manufacturer includes only trials post-2000 based on an argument of heterogeneity that to the ERG does not seem to be solved by excluding the pre-2000 trials. The ERG requested further MTC analyses to address the issue of heterogeneity in the points for clarification, but the manufacturer was unable to provide the necessary information in time.

Other issues raised by the ERG significantly influence the ICER and have been summarised in Section 6.3. The ERG considers the exclusion of non-health costs and use of full incremental probabilistic analysis necessary to meet the NICE reference case analysis. Decisions will need to be made on the most appropriate initial population, the most appropriate natural history data, the most appropriate estimates of HRQoL and whether to accept the ERG's recalculation of SPMS conversion.

8.1 Implications for research

A good quality placebo-controlled trial of teriflunomide powered for 6 month SAD as the primary outcome is required. Similar trials comparing teriflunomide with the beta-interferons, glatiramer acetate and new DMTs are also required.

A MTC including all relevant studies and controlling for heterogeneity is needed.

9 References

1. Pokorski RJ. Long-term survival experience of patients with multiple sclerosis. *J Insur Med* 1997;29:101-6.

2. Department of Health (DH). *Cost effective provision of disease modifying therapies for people with multiple sclerosis*; 2002. Available from: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4012214.pdf

3. Zajicek JP, Ingram WM, Vickery J, Creanor S, Wright DE, Hobart JC. Patient-orientated longitudinal study of multiple sclerosis in south west England (The South West Impact of Multiple Sclerosis Project, SWIMS) 1: protocol and baseline characteristics of cohort. *BMC Neurol* 2010;10:88.

4. Genzyme. *Meeting report: Teriflunomide Advisory Board (held 22nd November 2011)*; 2011

5. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). *CHMP summary of positive opinion for Aubagio. 21 March 2013 EMA/167896/2013* 2013. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/002514/WC500140690.pdf

6. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). *Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis. Draft. 20 September. EMA/CHMP/771815/2011.*; 2012. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/10/WC500133438.pdf

7. Centre for Reviews and Dissemination (CRD). *Systematic Reviews. CRD's guidance for undertaking reviews in health care.* York: University of York; 2009.

8. Whitehead A. *Meta-Analysis Of Controlled Clinical Trials*. Chichester (England): John Wiley & Sons, Ltd; 2003.

9. O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *N Engl J Med*;365:1293-303.

10. Pickin M, Cooper CL, Chater T, O'Hagan A, Abrams KR, Cooper NJ, et al. The Multiple Sclerosis Risk Sharing Scheme Monitoring Study--early results and lessons for the future. *BMC Neurol* 2009;9:1.

11. World Health Organization (WHO). *Atlas multiple sclerosis resources in the world 2008*. Geneva: WHO; 2008.

12. Office for National Statistics. *Ethnicity and National Identity in England and Wales 2011*. 2012. [cited 2013 15/07]. Available from: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rpt-ethnicity.html.

13. Miller AE, O'Connor P, Wolinsky JS, Confavreux C, Kappos L, Olsson TP, et al. Pre-specified subgroup analyses of a placebo-controlled phase III trial (TEMSO) of oral teriflunomide in relapsing multiple sclerosis. *Mult Scler* 2012;18:1625-32.

14. Galetta SL. The controlled high risk Avonex multiple sclerosis trial (CHAMPS Study). *J Neuroophthalmol* 2001;21:292-5.

15. Rojas JI, Romano M, Ciapponi A, Patrucco L, Cristiano E. Interferon Beta for primary progressive multiple sclerosis. *Cochrane Database Syst Rev* 2010:CD006643.

16. Oliver BJ, Kohli E, Kasper LH. Interferon therapy in relapsing-remitting multiple sclerosis: a systematic review and meta-analysis of the comparative trials. *J Neurol Sci* 2011;302:96-105.

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

18. 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.

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

20. Bornstein MB, Miller A, Slagle S, Weitzman M, Crystal H, Drexler E, et al. A pilot trial of Cop 1 in exacerbating-remitting multiple sclerosis. *N Engl J Med* 1987;317:408-14.

21. 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.

22. Comi G, Filippi M, Wolinsky JS. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging--measured disease activity and burden in patients with relapsing multiple sclerosis. European/Canadian Glatiramer Acetate Study Group. *Ann Neurol* 2001;49:290-7.

23. De Stefano N, Curtin F, Stubinski B, Blevins G, Drulovic J, Issard D, et al. Rapid benefits of a new formulation of subcutaneous interferon beta-1a in relapsing-remitting multiple sclerosis. *Mult Scler* 2010;16:888-92.

24. Cadavid D, Wolansky LJ, Skurnick J, Lincoln J, Cheriyan J, Szczepanowski K, et al. Efficacy of treatment of MS with IFNbeta-1b or glatiramer acetate by monthly brain MRI in the BECOME study. *Neurology* 2009;72:1976-83.

25. Lublin F, Cofield S, Cutter G, Conwit R, Narayana P, Nelson F, et al. The CombiRx Trial: A Multi-Center, Double-Blind, Randomized Study Comparing the Combined Use of Interferon Beta-1a and Glatiramer Acetate to Either Agent Alone in Participants with Relapsing-Remitting Multiple Sclerosis – Clinical Outcomes. *Neurology* 2012;78:PL02.003.

26. Vollmer T, Sorenson S. A placebo-controlled and active comparator phase III trial (BRAVO) for relapsing-remitting multiple sclerosis. *Mult Scler* 2011;17:S507-S08.

27. Mikol DD, Barkhof F, Chang P, Coyle PK, Jeffery DR, Schwid SR, et al. Comparison of subcutaneous interferon beta-1a with glatiramer acetate in patients with relapsing multiple sclerosis (the REbif vs Glatiramer Acetate in Relapsing MS Disease [REGARD] study): a multicentre, randomised, parallel, open-label trial. *Lancet Neurology* 2008;7:903-14.

28. Cohen JA, Barkhof F, Comi G, Hartung HP, Khatri BO, Montalban X, et al. Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *N Engl J Med* 2010;362:402-15.

29. Panitch H, Goodin DS, Francis G, Chang P, Coyle PK, O'Connor P, et al. Randomized, comparative study of interferon beta-1a treatment regimens in MS: The EVIDENCE Trial. *Neurology* 2002;59:1496-506.

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

31. Longworth L, Tosh J, Abrams K. *Lapatinib for the treatment of advanced and metastatic breast cancer: a review of the response to the ACD provided by the manufacturer of Lapatinib*: NICE Decision Support Unit; 2008. Available from: http://www.nicedsu.org.uk/PDFs%20of%20reports/Lapatinib%20DSU%20final%20report.pdf

32. National Institute for Health and Clinical Excellence (NICE). *Guide to the methods of technology appraisal 2013*. London: National Institute of Health and Care Excellence; 2013. Available from: http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9/the-reference-case#presentation-of-data-and-results

33. Trojano M, Pellegrini F, Fuiani A, Paolicelli D, Zipoli V, Zimatore GB, et al. New natural history of interferon-beta-treated relapsing multiple sclerosis. *Ann Neurol* 2007;61:300-6.

34. Held U, Heigenhauser L, Shang C, Kappos L, Polman C. Predictors of relapse rate in MS clinical trials. *Neurology* 2005;65:1769-73.

35. Patzold U, Pocklington PR. Course of multiple sclerosis. First results of a prospective study carried out of 102 MS patients from 1976-1980. *Acta Neurol Scand* 1982;65:248-66.

36. Orme M, Kerrigan J, Tyas D, Russell N, Nixon R. The effect of disease, functional status, and relapses on the utility of people with multiple sclerosis in the UK. *Value Health* 2007;10:54-60.

37. Dee A, Hutchinson M, De La Harpe D. A budget impact analysis of natalizumab use in Ireland. *Ir J Med Sci* 2012;181:199-204.

38. Gani R, Giovannoni G, Bates D, Kemball B, Hughes S, Kerrigan J. Cost-effectiveness analyses of natalizumab (Tysabri) compared with other disease-modifying therapies for people with highly active relapsing-remitting multiple sclerosis in the UK. *Pharmacoeconomics* 2008;26:617-27.

39. Prosser LA, Kuntz KM, Bar-Or A, Weinstein MC. Patient and community preferences for treatments and health states in multiple sclerosis. *Mult Scler* 2003;9:311-9.

40. Prosser LA, Kuntz KM, Bar-Or A, Weinstein MC. Cost-effectiveness of interferon beta-1a, interferon beta-1b, and glatiramer acetate in newly diagnosed non-primary progressive multiple sclerosis. *Value Health* 2004;7:554-68.

41. Boye KS, Matza LS, Walter KN, Van Brunt K, Palsgrove AC, Tynan A. Utilities and disutilities for attributes of injectable treatments for type 2 diabetes. *Eur J Health Econ* 2011;12:219-30.

42. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health and quality of life outcomes* 2008;6:84.

43. van Hoek AJ, Underwood A, Jit M, Miller E, Edmunds WJ. The impact of pandemic influenza H1N1 on health-related quality of life: a prospective population-based study. *PLoS One* 2011;6:e17030. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21399678

44. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making* 2006;26:410-20.

45. Freedman MS, Confavreux C, Comi G, Kappos L, Olsson TP, Miller A. *Hair thinning associated with teriflunomide therapy is manageable*. In: 17th Americas Comittee for Tretament and Research of Multiple Sclerosis (ACTRIMS). San Diego, California; 2012.

46. NHS Choices. *Bronchitis*. 2012. [cited 2013 7 May]. Available from: http://www.nhs.uk/conditions/Bronchitis/Pages/Introduction.aspx

47. Tyas D, Kerrigan J, Russell N, Nixon R. The distribution of the cost of multiple sclerosis in the UK: how do costs vary by illness severity? *Value Health* 2007;10:386-9.

48. Karampampa K, Gustavsson A, Miltenburger C, Tyas D. Treatment experience, burden and unmet needs (TRIBUNE) in MS study: results from the United Kingdom. *Mult Scler* 2012;18:41-5.

49. British National Forumulary (BNF). *Natalizumab drug information*. [cited 2013 22 May].

10 Appendices

Appendix Tables of comparison between direct and MTC results (base case) for comparisons included in the model

Appendix Table Comparison between direct and MTC results (Base case) for comparisons included in the model: ARR results

Comparison	Direct Evidence (Trial)	Result from MTC (Base case) (Random effects model)
Fingolimod (0.5mg) vs. placebo	Kappos L, 2010: 0.45 [0.37, 0.54] Saida, 2012: 0.51 [0.27, 0.95] Calabresi, 2012: 0.52 [0.43, 0.64] Direct MA: 0.486 [0.425, 0.556]	0.46 [0.4, 0.53]
Fingolimod (0.5mg) vs. Avonex 30µg 33	Cohen, 2010: 0.48 [0.36, 0.65]	[redacted]
GA 20 mg vs. placebo	CONFIRM: 0.72 [0.61, 0.87]	0.64 [0.53, 0.76]
GA 20 mg vs. Avonex 30µg 33	CombiRx: 0.69 [0.51, 0.93]	[redacted]
GA 20 mg vs. Rebif	Mikol, 2008: 0.97 [0.79, 1.18]	[redacted]
GA 20 mg vs. Betaferon 250	O'Connor P, 2009: 0.94 [0.84, 1.06] Cadavid, 2009: 0.89 [0.52, 1.53] Direct MA: 0.942 [0.841, 1.055]	[redacted]
Natalizumab (300mg) vs. placebo	Polman, 2006: 0.32 [0.27, 0.37]	0.31 [0.25, 0.39]
Teriflunomide 14 mg vs. placebo	O'Connor P, 2006: 0.68 [0.36, 1.27] TEMSO: 0.69 [0.58, 0.81] TOWER: 0.64 [0.53, 0.77] Direct MA: 0.664 [0.589, 0.749]	0.67 [0.57, 0.77]
Teriflunomide 14 mg vs. Rebif 44	TENERE: 1.2 [0.71, 2.01]	1.06 [0.84, 1.35]
Rebif 44 vs. placebo*		0.62 [0.51, 0.76]
Avonex 30µg vs. placebo*		0.78 [0.67, 0.91]
Betaferon vs. placebo*		0.68 [0.52, 0.87]

* No direct comparison trial for beta-interferons vs. placebo was included in MTC; GA glatiramer acetate

Appendix Table Comparison between direct and MTC results (Base case) for comparisons included in the model: 3 month SAD

Comparison	Direct Evidence (Trial)	Result from MTC (Random effects model)
Fingolimod (0.5mg)vs. placebo	Kappos L, 2010: 0.7 [0.52, 0.94] Calabresi, 2012: 0.83 [0.61, 1.13] Direct MA: 0.76 [0.614, 0.941]	0.75 [0.58, 0.96]
Fingolimod (0.5mg)vs. Avonex 30µg 33	Cohen, 2010: 0.74 [0.45, 1.23]	[redacted]
GA 20 mg vs. placebo	CONFIRM: 0.93 [0.63, 1.37]	0.93 [0.59, 1.45]
GA 20 mg vs. Betaferon 250	O'Connor P, 2009: 0.75 [0.56, 1.02]	[redacted]
Natalizumab (300mg) vs. placebo	Polman, 2006: 0.58 [0.43, 0.78]	0.58 [0.4, 0.84]
Teriflunomide 14 mg vs. placebo	TEMSO: 0.7 [0.51, 0.96] TOWER: 0.68 [0.47, 1] Direct MA: 0.694 [0.544, 0.886]	0.71 [0.53, 0.92]
Teriflunomide 14 mg vs. Rebif 44	TENERE: 0.98 [0.4, 2.42]	0.9 [0.54, 1.45]
Rebif 44 vs. placebo*	HR 0.65 (95% CI 0.45, 0.94)	HR 0.79 (95% CI 0.51, 1.24)
Rebif 44 vs. Avonex 30µg 33	Pan itch, 2002: 0.87 [0.58, 1.31]	[redacted]
Avonex 30µg vs. placebo*	NA**	HR 0.90 (95% CI 0.58, 1.37)
Betaferon vs. placebo*	HR 0.68 (95% CI 0.40, 1.17)	HR 1.21 (95% CI 0.68, 2.16)

* No direct comparison trial for beta-interferons vs. placebo included in MTC Sourced from Table AA2.3 of manufacturer's submission; ** Data only available for 6 mth SAD; GA glatiramer acetate

AppendixTable : 3 mth SAD results, comparison between direct and indirect result from MTC with direct evidence from placebo controlled trials not included in the MTC base case.

Comparison	**Direct evidence *(Trial) HR (95% CI)**	**Result from MTC base case HR (95% CI)**	**Result from MTC All years HR (95% CI)**
Rebif 44 vs. placebo	HR 0.65 (95% CI 0.45, 0.94)	HR 0.79 (95% CI 0.51, 1.24)	[redacted]
Avonex 30µg vs. placebo	NA**	HR 0.90 (95% CI 0.58, 1.37)	[redacted]
Betaferon vs. placebo	HR 0.68 (95% CI 0.40, 1.17)	HR 1.21 (95% CI 0.68, 2.16)	[redacted]

*Sourced from Table AA2.3 of manufacturer's submission; ** Data only available for 6 mth SAD.

Appendix Table Comparison between direct and MTC results (Base case) for comparisons included in the model: Discontinuations

Comparison	Direct Evidence (Trial)	Result from MTC (Random effects model)
Fingolimod (0.5mg)vs. placebo	Kappos L, 2010: 0.61 [0.44, 0.85] Saida, 2012: 1.59 [0.53, 4.82] Direct MA: 0.848 [0.348, 2.068]	[redacted]
Fingolimod (0.5mg)vs. Avonex 30µg 33	Cohen, 2010: 0.85 [0.56, 1.31]	[redacted]
GA 20 mg vs. placebo	CONFIRM: 0.59 [0.43, 0.81]	[redacted]
GA 20 mg vs. Avonex 30µg 33		[redacted]
GA 20 mg vs. Rebif	Mikol, 2008: 0.6 [0.41, 0.88]	[redacted]
GA 20 mg vs. Betaferon 250	O'Connor P, 2009: 1.43 [1.03, 1.98] Cadavid, 2009: 0.91 [0.21, 3.96] Direct MA: 1.401 [1.019, 1.926]	[redacted]
Natalizumab (300mg) vs. placebo	Polman, 2006: 0.81 [0.54, 1.2]	[redacted]
Teriflunomide 14 mg vs. placebo	O'Connor P, 2006: 3.8 [1.15, 12.58] TEMSO: 0.9 [0.65, 1.25] TOWER: 1.07 [0.79, 1.45] Direct MA: 1.337 [0.564, 3.17]	[redacted]
Teriflunomide 14 mg vs. Rebif 44	TENERE: 0.59 [0.31, 1.11]	[redacted]
Rebif 44 vs. placebo*		
Rebif vs. Avonex 30µg	Panitch, 2002: 1.2 [0.66, 2.19]	[redacted]
Avonex 30µg vs. placebo*		
Betaferon vs. placebo*		

* No direct comparison trial for beta-interferons vs. placebo included in MTC; GA glatiramer acetate