THE EFFECTIVENESS AND COST-EFFECTIVENESS OF NATALIZUMAB FOR MULTIPLE SCLEROSIS: AN EVIDENCE REVIEW OF THE SUBMISSION FROM BIOGEN

Report commissioned by: NHS R&D HTA Programme

On behalf of: NICE

Produced by: Peninsula Technology Assessment Group (PenTAG)
Peninsula Medical School,
Wessex Institute for Health Research and Development,
University of Southampton.

Authors:

[redacted] Research Fellow, Peninsula Technology Assessment Group
[redacted] Senior Lecturer, Peninsula Technology Assessment Group
[redacted] Research Fellow, Peninsula Technology Assessment Group
[redacted] Associate Professor, Peninsula Technology Assessment Group
[redacted] Information Scientist, Southampton Health Technology Assessment Group
[redacted] Senior Lecturer in Public Health, Peninsula Technology Assessment Group

Correspondence to: [redacted]

Date completed : 6 February 2007

ABOUT THE PENINSULA TECHNOLOGY ASSESSMENT GROUP (PENTAG)

The Peninsula Technology Assessment Group is part of the Institute of Health and Social Care Research at the Peninsula Medical School. PenTAG was established in 2000 and carries out independent Health Technology Assessments for the UK HTA Programme and other local and national decision-makers. The group is multi-disciplinary and draws on individuals' backgrounds in public health, health services research, computing and decision analysis, systematic reviewing, statistics and health economics. The Peninsula Medical School is a school within the Universities of Plymouth and Exeter. The Institute of Health and Social Care Research is made up of discrete but methodologically related research groups, among which Health Technology Assessment is a strong and recurring theme. Projects to date include:

- The Effectiveness And Cost-Effectiveness Of Imatinib (STI 571) In Chronic Myeloid Leukaemia - A Systematic Review (2002)
- Screening For Hepatitis C Among Injecting Drug Users And In Genitourinary Medicine (GUM) Clinics - Systematic Reviews Of Effectiveness, Modelling Study And National Survey Of Current Practice (2002)
- Systematic Review Of Endoscopic Sinus Surgery For Nasal Polyps (2003)
- The Effectiveness And Cost-Effectiveness Of Imatinib For First Line Treatment Of Chronic Myeloid Leukaemia In Chronic Phase (2003)
- The Effectiveness And Cost-Effectiveness Of Microwave And Thermal Balloon Endometrial Ablation For Heavy Menstrual Bleeding - A Systematic Review And Economic Modelling (2004)
- Do The Findings Of Case Series Studies Vary Significantly According To Methodological Characteristics?(2005)
- The Effectiveness And Cost-Effectiveness Of Pimecrolimus And Tacrolimus For Atopic Eczema - A Systematic Review And Economic Modelling (2005)
- The Effectiveness And Cost-effectiveness Of Dual Chamber Pacemakers Compared To Single Chamber Pacemakers For Bradycardia Due To Atrioventricular Block Or Sick Sinus Syndrome - Systematic Review And Economic Evaluation (2005)
- The Effectiveness And Cost-Effectiveness Of Surveillance Of Barrett's Oesophagus: Exploring The Uncertainty (2006, In Press)

- The Effectiveness And Cost-Effectiveness Of Carmustine Wafers And Temozolomide For Newly Diagnosed High Grade Glioma (2006, In Press)

Source of funding

This report was commissioned by the NHS R&D HTA programme.

Competing Interests of Authors

[redacted] is an applicant on a research project recently funded by the Multiple Sclerosis Society.

Other authors have no pecuniary relationship with companies making or profiting from treatments for multiple sclerosis.

Acknowledgements

We particularly acknowledge the help of the clinical experts who advised us – [redacted], Senior Lecturer of Clinical Neuroscience and Consultant Neurologist and [redacted], Professor of Clinical Neuroscience and Honorary Consultant Neurologist both at the Peninsula Medical School and Derriford Hospital Plymouth.

Any errors that remain are the responsibility of the authors.

With many thanks to [redacted](PenTAG) for his insightful comments on earlier drafts of this report.

With many thanks to [redacted]PenTAG) for administrative project support.

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

Contributions of Authors

[redacted]	Provided overall project management. Critiqued the effectiveness evaluation provided by industry. Contributed to the writing of the report.
[redacted]	Critiqued the economic evaluation provided by industry. Contributed to the writing of the report
[redacted]	Critiqued the model provided by industry. Contributed to the writing of the report.
[redacted]	Undertook literature searches for report, commented on the searches provided by industry.
[redacted]	Contributed to the critique of the industry submission and report writing.
[redacted]	Critiqued the statistical analyses provided by industry. Contributed to the writing of the report.

Table of contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

AE	Adverse effects
CEA	Cost-effectiveness analysis
CiC	Commercial in confidence
EDSS	Expanded Disability Status Scale
Gd	Gadolinium
GA	Glatiramer acetate
HARRMS	Highly active relapsing remitting multiple sclerosis
IFN- β	Beta-interferon
ITT	Intention to treat
MRI	Magnetic Resonance Imaging
MS	Multiple sclerosis
MSM	Multi-state model
MTX	Mitoxantrone
NAB	Neutralising anti-bodies
NAT	Natalizumab
OR	Odds ratio
PML	Progressive Multifocal Leukoencephalopathy
RES	Rapidly evolving severe
RR	Relative risk
RRMS	Relapsing remitting multiple sclerosis
SOT	Sub-optimal therapy
SPMS	Secondary progressive multiple sclerosis

1 SUMMARY

1.1 Scope of submission

- The effectiveness and cost-effectiveness of natalizumab for the treatment of those with highly active relapsing remitting multiple sclerosis compared to best supportive care, beta-interferon and glatiramer acetate.
- Highly active relapsing remitting multiple sclerosis includes two subgroups – those in whom relapses occur at least twice in one year (the rapidly evolving severe, RES group) and those who continue to have active disease despite treatment with beta-interferon (the sub optimal therapy, SOT group). These groups are in line with the licensed indications.

1.2 Summary of submitted clinical effectiveness evidence

- One RCT, the AFFIRM trial, comparing natalizumab with placebo in people with relapsing remitting multiple sclerosis, forms the basis of the submission. Subgroup analysis (n=209) provides information about those with rapidly evolving severe disease.
- Direct data about the sub-optimal therapy group is not available. The submission assumes effectiveness in this group based on an RCT of natalizumab in addition to beta-interferon compared with continued beta-interferon alone in people with active disease despite beta-interferon treatment. That is, the submission assumes that monotherapy compared to placebo will show a similar impact to combination therapy compared to sub-optimal treatment with beta-interferon.
- Results show that natalizumab is effective reducing sustained disability progression and the number of relapses compared to placebo in those with relapsing remitting multiple sclerosis and those with rapidly evolving severe multiple sclerosis.
- Indirect comparisons of effectiveness were undertaken as no head to head trials exist between natalizumab and active comparators beta-interferon and glatiramer acetate.

[redacted] Natalizumab is more effective at reducing relapses.

- No analysis of natalizumab compared to MTX is included.

1.3 Summary of submitted cost-effectiveness of evidence

- A decision analytic (Markov) model is used to estimate the costs and benefits of treatment with natalizumab and comparators.
- The model is based on disability progression (measured using the Expanded Disability Status Scale) and number of relapses.
- The model uses clinical effectiveness data for disability progression and relapse rate from the AFFIRM trial for natalizumab. This is supplemented by disability progression data from a large observational data set, the London-Ontario dataset.
- Clinical effectiveness data for disability progression and relapse rate for beta-interferon and glatiramer are taken form Cochrane reviews.
- A cross-sectional postal survey, the UK MS survey, is used to supply resource use and utility data for the model.
- In the rapidly evolving severe group, natalizumab showed a cost per quality-adjusted-life-year gain of £32,000, £35,000 and £45,000 compared with beta-interferon and glatiramer acetate and best supportive care. Corresponding estimates in the SOT group were £43,000, £44,000 and £56,000 per QALY.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

- The main RCT used in the submission was well conducted when assessed using the NICE internal validity criteria.
- The approach taken to model the disease is pragmatic given the available data and previous MS models.

1.4.2 Limitations

- There is no direct evidence about the use of natalizumab among the SOT group.
- Evidence for the RES group is based on a subgroup analysis of one RCT.
- There are no head to head comparisons of natalizumab with other active therapies.
- We are unsure about the appropriateness of some of the data used to populate the model for the patient group under consideration.
- Although frequently used, the Expanded Disability Status Scale, on which the model is based, has some well known limitations.

1.4.3 Areas of uncertainty

- The effect of natalizumab compared with active treatments is uncertain – indirect comparisons among people with highly active MS show wide confidence intervals, that include no benefit, around the key outcome of disease progression.
- Underlying disease progression in the model is based on data from the AFFIRM trial and should be treated with caution.

2 BACKGROUND

This report provides a review of the evidence submitted by Biogen in support of natalizumab for the treatment of people with highly active multiple sclerosis. It considers both the original submission received on 28 November 2006 and a subsequent addendum supplied by Biogen on 12 December 2006. The addendum was produced in response to our immediate queries relating to miscalculations in the economic model.

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

Background information about the condition of multiple sclerosis (MS) is spread throughout the submission. The beginning of Chapter 4 (p.32 of the submission) is headed as describing the condition but this is presented as a short bullet pointed list which, although it briefly outlines the impact of disease and its prevalence, does not provide details about aetiology, epidemiology, prognosis or symptoms. Details of the impact of MS on quality of life, especially as disability progresses, are provided in the submission as this is key to the modelling approach used.

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

The submission correctly notes that there is currently no standard treatment or guidelines in the UK for highly active relapsing remitting MS (HARRMS). A plausible treatment pathway for disease modifying treatments in RRMS, based on NICE guidelines, is illustrated (p.33 of the submission).[1] This includes the risk-sharing scheme for beta-interferon (IFN-β) and glatiramer acetate (GA). Details of the risk sharing scheme are not given, in particular regarding the specific initiation and stopping criteria for treatment with IFN- β or GA. Treatment of specific impairments, and treatment of acute episodes (relapses) are not considered although relapse frequency and severity are not central to the evaluation.

3 CRITIQUE OF THE MANUFACTURER'S DEFINITION OF THE DECISION PROBLEM

3.1 Population

In accordance with the scope and the license for natalizumab as monotherapy in MS, two populations are considered in the submission. Both are types of highly active relapsing remitting multiple sclerosis (HARRMS), a subgroup of RRMS:

1. Sub optimal treatment (SOT) subgroup.
2. Rapidly evolving severe (RES) subgroup.

People with RES have :

- Experienced >=2 disabling relapses in the prior year *and*
- Had >=1 gadolinium enhancing lesions on brain MRI *or* significant increase in T2 lesion load compared to a previous MRI

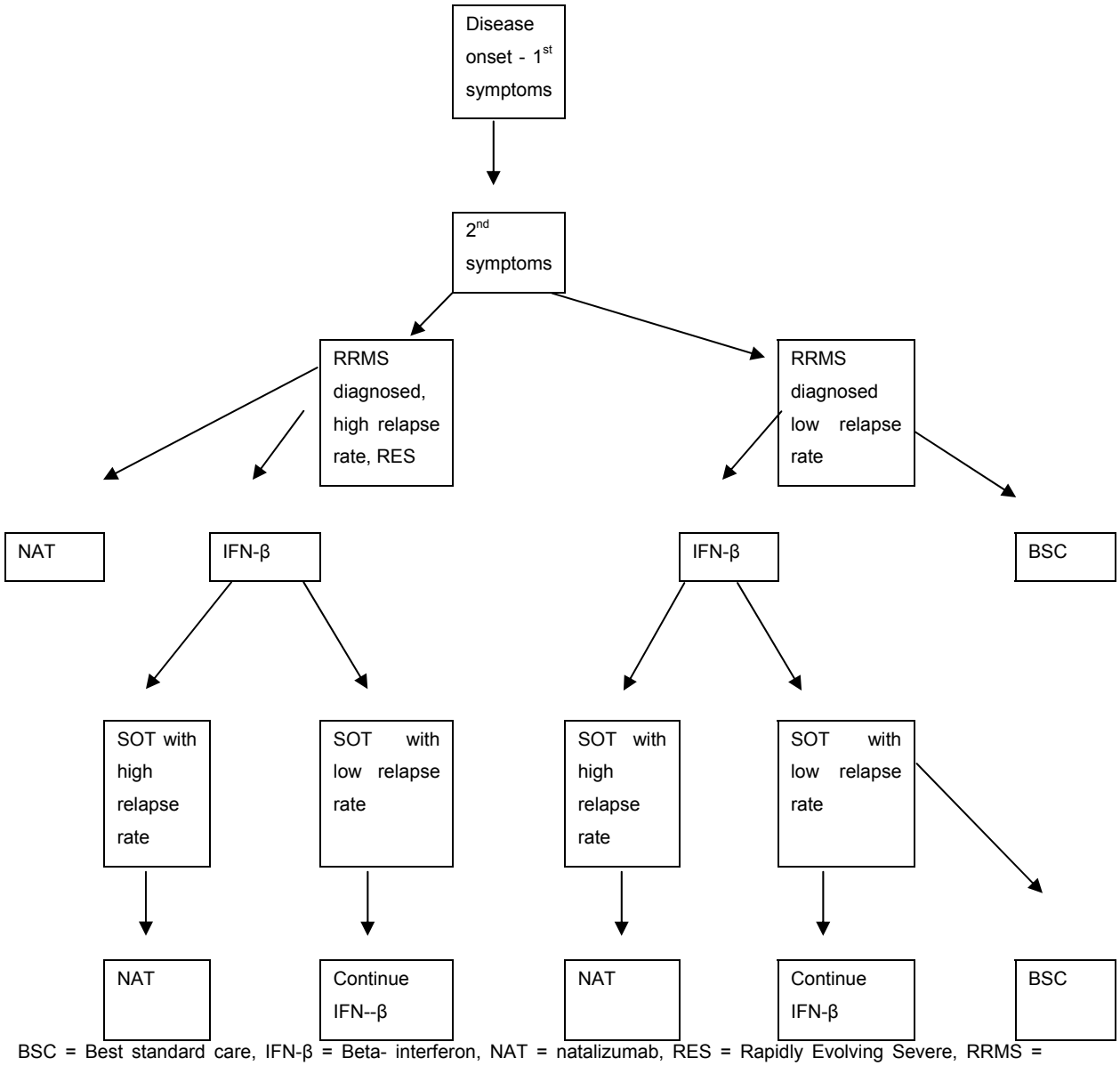
People with SOT have:

- Experienced >=1 disabling relapses in the previous year while receiving therapy *and*
- Had >=9 T2-hyperintense lesions on MRI *or* >=1 gadolinium enhancing lesions

These groups are subgroups of the larger HARRMS group although there may be some overlap. Possible disease developments and treatment (choices shown between natalizumab, IFN-β and best standard care (BSC) only for simplicity) are illustrated in Figure 1, where a diagnosis of RES is made if more than two relapses are experienced during a year. Severity, or duration of relapse does not form part of this definition. The diagnosis of MS itself may take some time to confirm. Its clinical definition is that two different areas of the central nervous system have been affected, and that these effects have been experienced on at least two separate occasions of at least one month apart. In addition, one or more objective neurological deficits should also be evident on examination and the person must be within the usual age range for the onset of MS.

A diagnosis of RES is therefore likely to relate to a group whose diagnosis of MS is made simultaneously with a diagnosis of RES as they present with obviously aggressive disease. Where this is not the case, not all patients will remain treatment naïve and currently, if given disease modifying treatment (DMT), are likely to receive IFBN-β. If they continue to experience relapses, then they will be considered as having SOT – this group may also be found to have RES if continued relapses are close together.

Figure 1: Possible disease development and place of treatment with natalizumab



BSC = Best standard care, IFN-β = Beta- interferon, NAT = natalizumab, RES = Rapidly Evolving Severe, RRMS = Relapsing remitting MS, SOT = Sub-optimal therapy

The SOT group may be the more likely clinical presentation relevant to treatment with natalizumab. The license for this group was granted based on the SENTINEL trial (described in more detail in Section 3.2 below).[2]

This trial compared the addition of natalizumab to beta-interferon (IFN-β) to IFN-β alone among a population who continued to experience relapses despite treatment. The basis for granting marketing license in this group in the European Medicine Agency's (EMEA) Public Assessment Report was:

> *Overall efficacy data suggest that efficacy in SENTINEL is mainly driven by natalizumab and not by Avonex [*IFN-β*] since Avonex by definition was not sufficiently active* (p.30 of the EMEA scientific discussion, quoted on p.22 of submission).

This statement therefore assumes that monotherapy with natalizumab is equivalent to combination therapy in this population. No supporting evidence for this assumption is provided.

It should be noted that neither the RES nor the SOT groups have formed the overall study population in an RCT of natalizumab monotherapy; 41% of participants recruited into the pivotal AFFIRM trial, which recruited people with RRMS, had experienced at least two relapses in the previous year and 49% had at least one gadolinium enhancing lesion.[3] A post-hoc analysis of a RES subgroup is provided as commercial in confidence material in the Biogen submission comprising 148 people treated with natalizumab and 61 treated with placebo.

Although the license for the SOT groups was based on results of the SENTINEL trial, the AFFIRM study is used as a proxy for the SOT group within the economic model. The Biogen submission suggests that the SOT group can be thought of as the same as the RRMS group, but at a later point in time (shown as pathway 3 in Figure 1). However, the AFFIRM population do not appear to have experienced extensive previous treatment, despite a median disease duration of five years.[4] Exclusion criteria for this trial include:

- Treatment with cyclophosphamide or mitoxantrone within the previous year.
- Treatment with IFN-β, GA, cyclosporine, azathioprine, methotrexate, or IV immune globule within the previous six months,
- Treatment for more than six months with IFN- β or GA.

The ITT population studied in the AFFIRM trial had an EDSS score of 0-6.0 with a mean of 2.3. Average age was 38 and ratio of women to men was 2.3:1. Median time since diagnosis was 5 years. The RES subgroup was similar in terms of EDSS, but slightly younger with an average age of 34.5 and with a higher ratio of women to men (3.5:1)

3.2 Intervention

The intervention described in the submission is natalizumab (Tysabri® manufactured by Biogen Idec Inc), a disease modifying agent that is first in a new drug class: selective adhesion molecule inhibitors. This matches the intervention outlined in the scope. Natalizumab is thought to reduce inflammation and demyelination by inhibiting migration of leukocytes across the blood brain barrier.

Natalizumab is administered as a 300mg dose by intravenous injection, once every four weeks in a hospital setting. The infusion is delivered over about one hour, at a rate of 2ml/minute. There are currently no restrictions on duration of treatment. Pharmacodynamic effects continue to be seen for about 12 weeks after the last dose.

As stated in the submission, natalizumab was granted marketing authorisation in the EU in July 2006.

3.3 Comparators

The NICE scope outlined the following comparators for this appraisal:

For adults with RRMS and high disease activity despite treatment with a beta-interferon (IFN-β) (SOT group):

- glatiramer acetate (GA) and mitoxantrone (MTX)
- standard care with no disease modifying treatment

For adults with rapidly evolving severe RRMS (RES group):

- IFB-β, GA and MTX
- standard care with no disease modifying treatment

The manufacturer considers that best supportive care (BSC, standard care) is not the most appropriate comparator for those with HARRMS. However, active disease modifying treatment (DMT) with IFN-β and GA was not recommended for use in England and Wales by NICE in 2002. Subsequent arrangements with manufacturers under a "risk sharing scheme" were set up by the Department of Health to allow selected people to receive these treatments. There have been no head-to-head trials to establish the effectiveness of natalizumab monotherapy compared with other disease modifying therapies (DMTs), and so placebo-controlled trials of each DMT are the only published evidence.

Based on meta-analysis of three randomised trials, IFN-β reduces the risk of relapse by about 30% at one year (RR0.73, 95% CI 0.55, 0.97; p= 0.03) and by 20% at two years (relative risk (RR) 0.80, 95% CI 0.73, 0.88; p<0.00001). Overall, MS progressed in 20% of those in the interferon arm and 29% of those in the placebo arm of these trials. Relapse reduction is equivalent to about one relapse avoided every 2.5 years of therapy.[5] IFN-β reduces disability progression by about 30% (RR 0.69, 95% CI 0.55, 0.87; p=0.002) compared to placebo over two years of treatment.[6]

Meta-analysis of two randomised GA trials in RRMS has shown a non-significant impact on the frequency of relapses and on disability progression compared to placebo.[7] At two years the RR of at least one relapse is 0.87 (95% CI 0.74, 1.02; p=0.08) and the RR of progression is 0.77 (95% CI 0.51, 1.14; p=0.2).

The Biogen submission rejects MTX as a valid comparator for natalizumab based on the fact that it is not currently licensed for MS treatment (p.29 of the submission). In addition, current NICE guidelines (2004) recommend its use only by those with experience of MTX, in an experimental setting where all risks are fully discussed with patients and with close monitoring for adverse effects.[1] The Biogen submission further includes data from the MS survey (2005, n=2048) showing that of the 288 people with RRMS in the UK who were taking DMT, none were taking MTX. These arguments seem reasonable. MTX carries risk of significant cardiac adverse effects and its use is restricted to a maximum of two years.[1]

Our clinical experts did note that MTX may occupy a similar clinical position in potential treatment strategies as natalizumab, while IFN-β is considered to be a milder treatment option (in terms of both effect and potential for serious adverse effects). MTX may be an appropriate comparator for natalizumab where IFN-β has failed, that is, in the SOT group. However, evidence for the effectiveness of MTX in the RES and SOT populations of interest is lacking. A Cochrane review of MTX for MS (2005) identified four placebo-controlled randomised trials involving 270 participants with RRMS, progressive relapsing MS (PRMS) and SPMS.[8] Only one study (n=65) reported on 24 week sustained disability at two-years with MTX, which would allow comparison with natalizumab trial outcomes here. This was in a population of secondary progressive and progressive relapsing MS (Odds ratio (OR) 0.3, 95% CI 0.09, 0.99 p=0.05). The same trial provides all available information in an analysis of annualised relapse rate at two years (WMD -0.85 (95% CI -0.47, -0.23, p=0.007). Other measures of relapse and disease progression based on trials in RRMS also suggest delays in progression and reduced relapse rates with MTX compared to placebo.

3.4 Outcomes specified in the systematic review

As there are no formal inclusion and exclusion criteria for trials discussed in the manufacturer's submission, no outcomes are specified as criteria for inclusion. The AFFIRM and SENTINEL RCTs both measure annualised relapse rate and cumulative progression rate. More details on these measures are described in section 4.1.6

3.5 Time frame

Natalizumab trial data provide outcomes at one and two years of follow up. However, MS is a long term, chronic condition. It is not known how rates of disease progression and risks of relapse reported in these trials should be extrapolated to the longer term nor how adverse effects might develop. It is known that natalizumab effectiveness decreases if persistent antibodies develop[9] although it is not yet known whether the incidence of antibodies will increase over time with natalizumab as it does with IFN-β.

4 CLINICAL EFFECTIVENESS

4.1 Critique of the manufacturer's approach

For a summary of the quality of the clinical effectiveness section of the Biogen submission on natalizumab see Appendix 1 on page93.

4.1.1 Description of search strategies and comment on whether the search strategies were appropriate

Clinical effectiveness searches for natalizumab

No formal systematic review of evidence to support the use of natalizumab in RES or SOT MS was undertaken for the submission. However, Biogen state that they are confident that they have identified all relevant information because of its recent license (June 2006 in Europe); the fact that any existing studies have been managed by the company (or under contact with them); and that Biogen has not authorised natalizumab supply to any third parties. This seems reasonable. However, our own search (details of search strategy in Appendix 2, page 98) showed a number of interim trial results – sometimes in abstract form and it may have been useful to state that these existed, and whether or not additional data was reported in them.[10-17] The trial described as MS231 in the submission text is not referenced to the publicly available paper and again it is not clear which data are in the public domain and which are from data files held by Biogen.[18] The trial described as MS201 does not appear to be available in a publicly published form but details were provided as CiC material in the Biogen submission.

No systematic searches were undertaken to identify evidence related to adverse effects with natalizumab in this submission.

Clinical effectiveness searches for comparators

The strategies used by the submission to identify relevant comparator trials were thorough (p. 192-219 of the submission). Cochrane systematic reviews by Rice and colleagues (2001) for IFN-β[6] and Munari and colleagues (2003) for GA[7] were updated for the submission by running literature searches for subsequently published evidence.

Detailed search strategies and results are provided, including the biomedical databases searched – Medline, Medline in Process, Embase and Cochrane Central, the time frame of the searches and host interface (Ovid) used. Suitable search terms were used with controlled language and text words. The searches are limited by RCT study and by year from 2001. The search strategies appear sound and reproducible.

We assessed whether or not there was additional trial data using the searches shown in Appendix 2. Fifty-three references were identified for GA and 305 for IFN-β through these update searches. After screening all abstracts and two full text papers, we did not identify any additional trials of IFN-β or GA that should have been included.

4.1.2 Statement of the inclusion and exclusion criteria and comment on whether they were appropriate

No explicit inclusion criteria were applied to the studies identified. Biogen have included in their submission RCT data about natalizumab in adults with RRMS and SPMS, as mono- or combination therapy, with follow up of at least 12 weeks, where natalizumab is taken as a monthly-dose course. Trials that describe single dose use of natalizumab during relapse are not included.[19]

There is only one completed RCT, the AFFIRM trial, in which people with RRMS received natalizumab as monotherapy compared to placebo and this trial provides most of the clinical data presented in the submission. However, it should be noted that the studied population were not failing on treatment (SOT), but could be considered treatment naïve (see Section 3.1). The RES group (n=209) is described in a post-hoc, CiC subgroup analysis of AFFIRM in the submission. The RES definition comprises three parts – number of relapses, and Gd-enhancing lesions or T2 lesions, each of these was separately specified as a subgroup for analyses in the statistical analyses plan for the AFFIRM study although the actual RES group itself was not. This post-hoc analysis was originally requested by the European Medicine Agency (EMEA).

In licensing natalizumab for the SOT subgroup, the EMEA based their decision on the SENTINEL trial, and this appears to be the main reason for including details of this RCT in the submission.

This trial examined a RRMS population who continued to experience relapses despite treatment with IFN-β, and showed that the addition of natalizumab reduced relapses and disease progression compared with IFN-β alone.[20] The EMEA concluded that the efficacy demonstrated by the treatment arm of SENTINEL was mainly driven by natalizumab, and was sufficient to allow recommendation of natalizumab for those in whom treatment with IFN-β failed. Natalizumab as a combination therapy is not, however, recommended.

4.1.3 Table of Identified Studies. What studies were included in the submission and what were excluded?

The four trials included in the submission, and their outcome measures, are shown in Table 1. See Appendix 3 on page 108 for a quality assessment of these trials. The trials appear to have been well conducted.

Table 1: Included natalizumab trials (taken from table 11 p. 57 of submission)

Study	Treatment group (n)	Type of patients	Outcomes
AFFIRM (Phase III registration study) RCT Once monthly (every 28 days) IV infusions. Two-year study.[21]	Natalizumab 300 mg by IV infusion every 4 weeks (n = 627) Placebo (n = 315)	Adults with RRMS.	*Primary:* Reduction in the rate of clinical relapses at one year Rate of sustained progression of disability at two years (EDSS) *Secondary:* Multiple MRI, progression, relapse and safety outcomes
SENTINEL (Phase III registration study) RCT Once monthly (every 28 days) IV infusions, adjunctive to IFN-beta. Two-year study.[22]	Natalizumab 300mg by IV infusion every 4 weeks interferon β-1a (n = 589) Interferon β-1a (n = 582)	Adults with RRMS.	*Primary:* Reduction in the rate of clinical relapses at one year Rate of sustained progression of disability at two years (EDSS) *Secondary:* Multiple MRI, progression, relapse and safety outcomes
MS 201 (Phase II) RCT Once monthly (every 28 days) IV infusions. Twelve-week study.	Natalizumab 3.0 mg/kg (n = 37) Placebo (n = 35)	Adults with RRMS or SPMS.	*Primary:* The number of new active lesions during the 12 weeks following the first treatment assessed by MRI. *Secondary:* Multiple MRI, progression, relapse and safety outcomes
MS 231 (Phase II) RCT Once monthly (every 28 days) IV infusions. Six-month study.[23]	Natalizumab 3.0 mg/kg (n = 68) Natalizumab 6.0 mg/kg (n = 74) Placebo (n = 71)	Adults with RRMS or SPMS.	*Primary*: Brain lesion activity assessed by magnetic resonance imaging (MRI) *Secondary*: Multiple MRI, progression, relapse and safety outcomes

Main trial: AFFIRM

One trial is central to the submission; the multicentre AFFIRM RCT comprising 942 people with RRMS (627 received natalizumab, 315 received placebo). Included were people aged 18-50 who had:

- A diagnosis of relapsing MS as defined by the McDonald criteria
- A baseline EDSS score of between 0.0 and 5.0,
- A brain scan demonstrating lesion(s) consistent with MS.

Data on a subset of people with RES are provided in the submission as CiC data. This subgroup is defined as those experiencing two or more relapses in the previous year, and having at least one lesion on gadolinium (Gd)-enhancing MRI or a significant increase in T_2 lesion load compared to a previous, recent MRI. These three groups were specified separately as subgroups in the original statistical analysis plan although RES itself is, strictly, a subgroup defined *post-hoc*. Of those with two or more relapses in the previous year, most had two relapses (n=299, with 83 people having more than two relapses). Biogen state that subgroup analyses for number of relapses, number of Gd-enhancing lesions and number of T2 lesions were all statistically significant.

[redacted]

The RES group comprised 209 people (148 received natalizumab and 61 placebo). The treatment group were younger than the placebo group (33.7yrs ±8.4 vs 36.4 yrs ±8.1, p=0.037). There was a non-significant difference in disease duration, with those treated with natalizumab diagnosed for a median of 4.0 years compared to 5.0 years for those treated by placebo. (p=0.531) These differences are not likely to have a substantial clinical impact and no adjustment was made to the statistical analysis.

Subsidiary trials

In addition to the AFFIRM trial, three studies provide additional information. (See Appendix 3 on page 108 for a quality assessment of these trials.) The SENTINEL trial compared the impact of natalizumab in addition to IFN-β to IFN-β alone, in patients with active RRMS despite treatment (i.e. the SOT group).

A total of 1171 people were included. This data is reported as supporting evidence for a reduction in relapses and disease progression with natalizumab although natalizumab is not licensed for this combination therapy use.

In addition, two smaller trials, MS231 (n=213) and MS201 (n=72) were also considered. Follow up in these trials was for 24 and 12 weeks respectively so, appropriately, these do not contribute to the data about annualised relapse rate and disease progression. However, safety data from these trials is used.

Comparators

Cochrane reviews were used as the basis for evidence about the effectiveness of the two active comparators, IFN-β and GA. A quality assessment of these systematic reviews is shown in Appendix 4 (page 115). Searches for these reviews were updated by the manufacturer, using a more limited population than the original in order to restrict trials to adult RRMS populations with relevant outcomes. No additional trials were identified for inclusion in the IFN-β review.

One additional study was identified for the GA review which assessed orally administered GA compared to placebo in 1651 adults with RRMS.[24] Outcomes were number of relapses, number of lesions and adverse effects. Follow up was for 14 months and relapse rates were not reported in ways suitable for inclusion in the existing meta-analysis. This paper concluded that there was no evidence of oral GA affecting relapse rate or MRI measures. Appropriately, this paper was not included in the meta-analysis of studies examining subcutaneous GA which is appropriate.

4.1.4 Details of any relevant studies that were not included in the submission

As there are no formal inclusion and exclusion criteria in the submission, it is difficult to say whether or not other studies should have been included. We did not identify any additional RCTs of natalizumab as monotherapy (the licensed indication). However, the SENTINEL trial, of natalizumab in combination with IFN-β is described in the submission. We identified one other combination therapy trial: the glatiramer acetate and natalizumab combination (GLANCE) trial.[25]

This may also represent a SOT group because all participants had experienced at least one relapse over the previous year, having been treated with GA for at least 12 months prior to randomisation. Although follow up was short, two monotherapy trials, MS201 and MS231 with short follow up did contribute adverse effect data to the submission.

GLANCE is a double-blind, randomised, placebo-controlled, parallel-group safety study with 110 subjects assessed safety and efficacy. All participants received the standard dose of GA (20 mg/day) together with 300 mg of natalizumab or placebo by intravenous infusion every four weeks. Follow up was up to 24 weeks.

All patients had a definite diagnosis of RRMS, were aged between of 18 and 55, had a baseline EDSS score between 0 and 5. The primary safety outcome was the rate of development of new active lesions on cranial MRI scans during a six-month period. Safety assessments included the incidence and severity of adverse events, and additional end points included EDSS score and number of relapses, gadolinium-enhancing lesions, T1 lesions, and T2 lesions.

The number of new T2 lesions was reduced 62% by using natalizumab, and the number of new gadolinium-enhancing lesions was reduced by 74%. The annualized relapse rate was reduced by 40%, although this was not statistically significant. There were no hypersensitivity reactions during the time of infusions. Regarding immunogenicity, the incidence of persistent positive antibodies in this study in the combination group was 13%.

Ongoing studies

Biogen reports on three ongoing studies relevant to the submission. One is an open label extension of the AFFIRM and SENTINEL trials. Two are prospective, observational cohort studies. Safety is the primary interest of these studies. PenTAG asked for any interim results to be provided, but Biogen stated that none were available. We did not identify any additional ongoing trials of natalizumab that were not mentioned in the submission.

4.1.5 Description and critique of the manufacturer's approach to validity assessment

The manufacturer uses items based on the CONSORT statement to critically appraise the RCTs of both natalizumab and comparators. They also provide a Jadad score (p.229-252 of the submission). The submission concludes that the trials of natalizumab are as good or better than those for the comparators although, as they themselves note, as they have access to trial information on their own databases, this is perhaps not surprising – all assessment of comparators were based on the published trial reports only.

The conclusion that the natalizumab trials are well conducted is reasonable (see our assessment of quality in Appendix 3 p. 108). Methods of randomisation and blinding were adequate, and although there was some drop out, this was small and similar in both arms (8% natalizumab vs 10% placebo in the AFFIRM trial). However, little consideration is given to relevant external validity. The assessment concludes that epidemiology of MS in the multicentre trial is likely to be similar to that in the UK but do not note that the studied population does not match the licensed populations.

4.1.6 Description and critique of the manufacturer's selection of outcomes

Participants in the trial were assessed at scheduled clinic visits every 12 weeks. Primary outcomes considered in the main AFFIRM trial and the Biogen submission are appropriate and largely consistent with EMEA recommendations for studies in MS[26]:

- Clinical relapse rate at one year
- Cumulative disability progression at two years

The EMEA recommends that disability progression using Kurtzke's Expanded Disability Status Scale (EDSS) should be defined as that sustained over at least 24 weeks to minimise the potential impact of EDSS shortcomings in terms of reliability. While the primary outcome reported in the published trial is disability progression sustained over 12 weeks, details of 24-week sustained disability progression are also presented in the submission. Detail about how this measure is implemented in practice is not provided.

The Kaplan-Meier plot for progression shows some of the cohort progressing at 12 weeks – although it is not clear how they could be defined as having sustained progression for at least 12 weeks by this point in the trial (Figure 2 in Polman and colleagues, 2006[27]). Kaplan-Meier plot for progression defined as sustained by 24 weeks is not provided, but again detail of how this was measured in practice is not provided.

A clinical relapse was defined as new or recurrent neurological symptoms (not associated with fever or infection) lasting at least 24 hours and that were associated with new neurological signs. The onset of new symptoms prompted unscheduled visits to the treating neurologist with 72 hours, who referred the patients to an examining neurologist within five days if a relapse was suspected.

Annualised relapse rate is provided in the trials. This in line with EMEA guidance which also recommends that impact on relapse rate should be assessed over at least two years: this is not presented in the published paper but is supplied in the submission.[28] The published paper additionally supplies information about the number of relapse free patients at one and two years and these data are used in the submissions indirect comparisons of natalizumab with other DMT. The AFFIRM publication does not report on relapse severity, although the Biogen submission provides surrogate markers of severity – showing that 71% vs 63% of relapses for those treated with placebo and natalizumab respectively resulted in steroid treatment ($p<0.001$) while 9.7% and 3.4% respectively required hospitalisation ($p<0.001$) (p.75 of the submission). Among those experiencing a relapse, post-hoc analysis showed a significantly higher study end physical component score on the SF-36 for those treated with natalizumab. The impact of other DMTs on relapse severity is not considered in the submission.

Disease progression was measured using the ordinal EDSS, (see Table 2) which is the most widely used disability scale for MS trials, although it demonstrates some well-known limitations.[29;30] Limited responsiveness, validity, and inter- and intra-rater reliability have been demonstrated.[31]

Table 2: Kurtzke Expanded Disability Status Scale

0.0	Normal neurological examination
1.0	No disability, minimal signs in one FS
1.5	No disability, minimal signs in more than one FS
2.0	Minimal disability in one FS
2.5	Mild disability in one FS or minimal disability in two FS
3.0	Moderate disability in one FS, or mild disability in three or four FS. Fully ambulatory
3.5	Fully ambulatory but with moderate disability in one FS and more than minimal disability in several others
4.0	Fully ambulatory without aid, self-sufficient, up and about some 12 hours a day despite relatively severe disability; able to walk without aid or rest some 500 meters
4.5	Fully ambulatory without aid, up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance; characterized by relatively severe disability; able to walk without aid or rest some 300 meters.
5.0	Ambulatory without aid or rest for about 200 meters; disability severe enough to impair full daily activities (work a full day without special provisions)
5.5	Ambulatory without aid or rest for about 100 meters; disability severe enough to preclude full daily activities
6.0	Intermittent or unilateral constant assistance (cane, crutch, brace) required to walk about 100 meters with or without resting
6.5	Constant bilateral assistance (canes, crutches, braces) required to walk about 20 meters without resting
7.0	Unable to walk beyond approximately five meters even with aid, essentially restricted to wheelchair; wheels self in standard wheelchair and transfers alone; up and about in wheelchair some 12 hours a day
7.5	Unable to take more than a few steps; restricted to wheelchair; may need aid in transfer; wheels self but cannot carry on in standard wheelchair a full day; May require motorized wheelchair
8.0	Essentially restricted to bed or chair or perambulated in wheelchair, but may be out of bed itself much of the day; retains many self-care functions; generally has effective use of arms
8.5	Essentially restricted to bed much of day; has some effective use of arms retains some self care functions
9.0	Confined to bed; can still communicate and eat.
9.5	Totally helpless bed patient; unable to communicate effectively or eat/swallow
10.0	Death due to MS

FS = Functional system. Eight FS are considered in the assessment of EDSS pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral, other

EDSS is assessed by the clinician. Lower grades of disability relate to impairment, while higher grades concentrate on mobility.[31] It does not adequately assess upper limb function or cognitive impairment.[32] Distribution is typically bi-model with fewer people in the middle range states (notably 4 and 5) than less disabled states 1-2 and more disabled states 6-7.

Most people with RRMS progress slowly, and one study of RRMS found that only 43% of the cohort progressed to state 6 or above (needing a cane to walk with or worse disability) over 25 years.[33] Despite its problems, the EDSS remains the most used measure ands there is no consensus about which, if any, alternative measure would have been more appropriate.

Similar detail on progression of the RES subgroup is not known. In the AFFIRM and SENTINEL trials, disability progression was considered "sustained" if over two visits (12 weeks apart) an increase of 1.0 or more (from a baseline of 1.0 or more) was seen. For those with a baseline EDSS score of 0, progression was defined as an EDSS increase of 1.5 or more. A movement of at least 1.0 is an indicator of clinically meaningful progression and this is appropriate, especially at lower grades.[31] As noted above, the published trials only provide data on disability progression sustained for 12 weeks, but the submission provides data for the more stringent measure of disability sustained for 24 weeks although there is some ambiguity about how this was measured.

Secondary outcomes at year one were:

- Reduction in new or enlarging T2 hyperintense lesions on MRI scans
- Reduction in number of Gd-enhancing lesions on brain MRI scans
- Proportion of relapse free participants

Secondary outcomes at year two were:

- Reduction in rate of clinical relapse
- Attenuation of increase in T2 hyperintense lesions on MRI scans
- Attenuation of increase in T1 hyperintense lesions on MRI scans
- Slowing the progression of disability (as measure on the MSFC
- Sustaining quality of life (measured by MS QoL inventory/ SF-36)

We consider these to be appropriate.

No SF-36 data is reported in the published trial paper, and minimal data are supplied in the submission. This gives only the mean change in overall mental health component and physical health component. No baseline is supplied.

Data on adverse events was also collected and binding antibodies against natalizumab were assessed using an enzyme linked immunosorbent assay.

4.1.7 Description and critique of the manufacturer's statistical approach

The statistical analysis of the AFFIRM trial was well reported.

The AFFIRM trial was adequately powered. The sample size was based on 90% power to detect an assumed annualised relapse rate of 0.6 with natalizumab and 0.9 with placebo with 15% drop out and required 900 people at 5% significance. Progression rates at the end of two years were assumed to be 35% with placebo and 23% with natalizumab. These effect sizes were met in the trial.

About 9% of participants withdrew from the AFFIRM trial (8% natalizumab group, 10% placebo arm). Appropriately, ITT analyses were used to assess efficacy. Adverse effects analyses excluded three people who did not receive placebo as they were assigned. Full information about reasons for withdrawal is given and a flow chart provided.

Appropriate statistical methods were used to compare the two groups. The FDA statistical review did raise the issue that since randomisation was stratified by site, site should have been incorporated as a covariate in the primary analysis. However, such adjustment is not statistically mandatory and as the FDA statistical reviewers concede, sparse data in some sites may have made such adjustment for centres problematic.

The RES subgroup is based on a smaller subgroup of the AFFIRM trials (n=209). As the impact of natalizumab is greater than predicted, significant results are seen in this subgroup. Details of withdrawals are not given.

See page 31 below for the approach to indirect comparisons.

4.1.8 Summary statement about the review of clinical effectiveness

Despite the lack of an explicit and exhaustive systematic review, we do not believe that data about the efficacy of natalizumab in modifying HARRMS have been missed. However, the evidence base is very limited, with only one RCT providing data on disease progression when using natalizumab as monotherapy with follow-up of two years. In addition, this trial population does not reflect the licensed population. A post-hoc subgroup analysis based on only 209 people with RES MS (61 treated with placebo) is the sole source of such relevant data. It is not clear from where evidence for the SOT group should be derived. This population has not been explicitly studied in any trial. The license decision made an assumption that effectiveness when combined with IFN-β would be the same as with monotherapy in this group. Most of the submission assumes that the SOT group will respond in a similar way to the treatment naïve RRMS group examined in AFFIRM.

4.2 Summary of submitted evidence

As described in section 4.1.3, most data for the submission comes from two RCTs: the SENTINEL and the AFFIRM studies (total N=2133, 1216 taking natalizumab). The SENTINEL trial is given little prominence in the submission. SENTINEL assessed the impact of natalizumab combined with IFN-β compared to continued IFN-β in patients who continued to have relapses on this treatment. It thus examined the SOT population but did not investigate monotherapy natalizumab, and did not compare it to placebo. SENTINEL was stopped approximately one month early due to reports of two cases of progressive multifocal leukoencephalopathy (PML), one of which was fatal. Additional data on safety comes from two smaller trials (total n=287) with shorter follow up.

Details of the patient characteristics for both the ITT and the RES subgroup in the AFFIRM study, and the SENTINEL study are shown in Table 3. P-values refer to differences between the arms of each trial.

The Biogen submission suggests that the SOT group may be regarded as the same as the RRMS group but at a later stage in the disease. The median disease duration in the

SENTINEL trial is greater at seven years (range 1-34) compared to five years in the AFFIRM trial (range 1-34).

The AFFIRM trial includes people with similar numbers of relapses in the past year and similar mean EDSS score to the SENTINEL trial (see Table 3).

Table 3: Main patient characteristics of the AFFIRM and SENTINEL trials of natalizumab

	AFFIRM ITT				SENTINEL				AFFIRM RES subgroup			
	NAT(n = 627)	Placebo (n = 315)	Total (n = 942)	P-value	NAT+ IFN (n = 589)	IFN alone (n = 582)	Total (n = 1171)	P-value	NAT(n = 148)	Placebo (n = 61)	Total (n = 209)	P-value
Age — Years												
Mean	35.6 ± 8.5	36.7 ± 7.8	36.0 ± 8.3	0.056	38.8 ± 7.7	39.1 ±7.6	38.9 ±7.7	*	33.7 ± 8.4	36.4 ± 8.1	34.5 ± 8.4	0.037
Range	18–50	19–50	18–50		18-55	19-55	18-55					
Sex — no. of patients (%)												
Male	178 (28)	104 (33)	282 (30)	0.144	147 (25)	162 (28)	309 (26)	*	37 (25)	10 (16)	47 (22)	0.175
Female	449 (72)	211 (67)	660 (70)		442 (75)	420 (72)	862 (74)		111 (75)	51 (84)	162 (78)	
Race — no. of patients (%)												
White	603 (96)	296 (94)	899 (95)	0.126	550 (94)	542 (93)	1092 (93)	*	59 (97)	200 (96)	>0.999	59 (97)
Other	24 (4)	19 (6)	43 (5)		39 (7)	40 (7)	79 (7)		2 (3)	9 (4)		2 (3)
Disease duration — yr												
Median	5.0	6.0	5.0	0.511	7.0	8.0	7.0	0.02	4.0	5.0	5.0	0.501
Range	0-34	0-33	0-34		1-34	1-34	1-34		0-26	1-31	0-31	
No. of relapses in past yr — no. of patients (%)												
0	6 (< 1)	6 (2)	12 (1)		0	1 (<1)	1 (<1)		0	0	0	0.166
1	368 (59)	180 (57)	548 (58)		390 (66)	357 (61)	747 (64)		0	0	0	
2	197 (31)	102 (32)	299 (32)		153 (26)	174 (30)	327 (28)		110 (74)	47 (77)	157 (75)	
≥ 3	56 (9)	27 (9)	83 (9)		44 (7)	50 (9)	94 (8)		38 (26)	14 (23)	52 (25)	
Mean	1.53 ± 0.91	1.50 ± 0.77	1.52 ± 0.86	0.640	1.44 ±0.75	1.49 ±0.72	1.47 ±0.73	*				
Range	0–12	0–5	0–12		1-7	0-5	0-7					
EDSS score — no. of patients (%)												
0	31 (5)	18 (6)	49 (5)		24 (4)	19 (3)	43 (4)		7 (5)	4 (7)	11 (5)	0.389
1.0–1.5	179 (29)	94 (30)	273 (29)		145 (25)	143 (25)	288 (25)		38 (26)	17 (28)	55 (26)	
2.0–2.5	208 (33)	103 (33)	311 (33)		214 (36)	203 (35)	417 (36)		54 (36)	21 (34)	75 (36)	
3.0–3.5	130 (21)	63 (20)	193 (20)		125 (21)	126 (22)	251 (21)		29 (20)	13 (21)	42 (20)	

	AFFIRM ITT				SENTINEL				AFFIRM RES subgroup			
	NAT(n = 627)	Placebo (n = 315)	Total (n = 942)	P-value	NAT+ IFN (n = 589)	IFN alone (n = 582)	Total (n = 1171)	P-value	NAT(n = 148)	Placebo (n = 61)	Total (n = 209)	P-value
4.0–4.5	60 (10)	28 (9)	88 (9)		68 (12)	72 (12)	140 (12)		18 (12)	6 (10)	24 (11)	
5.0	17 (3)	7 (2)	24 (3)		12 (2)	16 (3)	28 (2)		2 (1)	0	2 (<1)	
≥ 5.5	2 (<1)	2 (<1)	4 (<1)		1 (<1)	3 (<1)	4 (<1)		0	0	0	
Mean	2.3 ± 1.2	2.3 ± 1.2	2.3 ± 1.2	0.784	2.4 ±1.1	2.5 ±1.1	2.4 ±1.1	*	2.4 ± 1.1	2.2 ± 1.1	2.3 ± 1.1	
Range	0–6	0–6	0–6		0-6.0	0-5.5	0-6.0		0-5	0-4.5	0-5	
No. of lesions on gadolinium-enhanced MRI — no. of patients (%)												
0	307 (49)	170 (54)	477 (51)		392 (67)	374 (64)	766 (65)		0	0	0	0.891
1	115 (18)	55 (17)	170 (18)		98 (17)	105 (18)	203 (17)		49 (33)	19 (31)	68 (33)	
2	66 (11)	24 (8)	90 (10)		31 (5)	32 (5)	63 (5)		25 (17)	14 (23)	39 (19)	
3	38 (6)	18 (6)	56 (6)		20 (3)	26 (4)	46 (4)		14 (9)	7 (11)	21 (10)	
≥ 4	100 (16)	46 (15)	146 (15)		43 (7)	42 (7)	85 (7)		60 (41)	21 (34)	81 (39)	
Missing data	1 (< 1)	2 (<1)	3 (<1)		5 (<1)	3 (<1)	8 (<1)		0	0	0	
Mean	2.2 ± 4.7	2.0 ± 4.8	2.2 ± 4.7		0.9 ±2.5	0.9 ±1.9	0.9 ±2.2		5.3 ± 6.3	5.4 ± 7.8	5.3 ± 6.8	
Range	0–36	0–39	0–39	0.511	0-24	0-24	0-24	*	1-34	1-39	1-39	
No. of lesions on T2-weighted MRI — no. of patients (%)												
<9	29 (5)	15 (5)	44 (5)	0.921	67 (11)	52 (9)	119 (10)	*	3 (2)	1 (2)	4 (2)	>0.999
≥9	597 (95)	299 (95)	896 (95)		519 (88)	528 (91)	1047 (89)		145 (98)	60 (98)	205 (98)	

4.2.1 Summary of results

Key results from the AFFIRM and SENTINEL studies are shown in Table 4. These show a statistically significant impact on both relapse rate and disease progression compared to placebo. The effect of natalizumab was consistent across subgroups regardless of age, gender, race, weight, baseline disease activity and MS disease history.

ITT analysis of AFFIRM

The AFFIRM trial showed a reduced risk of sustained (>12 weeks) progression at two years with 17% in the natalizumab arm progressing compared to 29% in the placebo arm (HR 0.58, 95% CI 0.43, 0.77; p<0.001). Using >24 weeks as the definition of sustained progression showed a relative risk reduction of 54% at two years (HR 0.46, 95% CI 0.33, 0.64; p<0.001). The Scottish Medicines Consortium (SMC) note in their summary statement that there is a small difference in mean EDSS score between the two groups.[34] At baseline, the mean EDSS in both groups was 2.3, and after 2 years this was increased by 0.04 (+/-0.86) for those receiving natalizumab and 0.41 (+/-1.09) for those receiving placebo. SMC note that the clinical meaning of a mean difference of 0.37 EDSS points is unclear. This data is not provided in the Biogen submission, but is available in a conference abstract.[35]

Relapses were also reduced with an annualised rate of relapse at two years of 0.24 with natalizumab compared with 0.73 with placebo (p<0.001). Data reproduced from the submission in Table 4 reports a relative risk *reduction* in relapses with natalizumab - relative risk at two years is 0.32 (95% CI 0.40 to 0.26). The clinical impact of reducing relapses by such an amount is not known. The AFFIRM trial suggests that without natalizumab treatment, someone with MS would experience one additional relapse over 16-18 months.[36] In addition, it should be noted that the relationship between a reduction in relapse rate and future disability progression is unclear.

A substantial improvement in lesion development was also seen in the AFFIRM trial. 96% of those treated with natalizumab showed no lesions on Gd-enhancement at one year. This compared with 53% having no lesions if treated with placebo (p<0.001).

Similarly no new T2 hyperintense lesions developed in 60% of those treated with natalizumab compared to 22% of those treated with placebo (p<0.001).

Efficacy was reported to be lost in the 6% of people who developed persistent antibodies to natalizumab, and they also experienced increased infusion related adverse effects. Further detail is not provided.

Limited detail is supplied about the impact of natalizumab on quality of life. No significant effect was seen on the MS Quality of Life Inventory (MSQLI). However, using the SF-36 showed small but statistically significant differences were seen on the summary domains for mental and physical health. The mental health summary is made up of vitality, role-emotional, mental health and social functioning dimensions (14 items in all) while the physical health summary is made up of physical functioning, role-physical, general health and bodily pain dimensions (21 items in all). Mean SF-36 at baseline was 2.3. With natalizumab, the mean mental component score increased by 2.00 (sd 10.91) compared to a reduction of 0.53 (sd 10.52) with placebo (p=0.011). The mean physical component score increased by 0.67 (sd 8.05) for those treated with natalizumab compared to reducing by -1.34 (sd8.47) with placebo (p=0.003). The absolute changes and differences are small and the value of this impact to people with MS is not clear.

RES subgroup analysis of AFFIRM

For the RES subgroup, similar reductions were seen to those in the ITT population. Cumulative probability of progression sustained for at least 12 weeks at two years was seen in 14% of those treated with natalizumab compared with 29% of those receiving placebo (HR 0.47, 95% CI 0.24, 0.93, p=0.029). For disability progression sustained over 24 weeks 10% of the natalizumab group progressed compared to 26% of the placebo group (HR 0.36, 95% CI 0.17, 0.76, p=0.008). Annualised relapse rates were 0.28 in the natalizumab group and 1.46 for those receiving placebo giving a relative risk reduction of 0.81 with natalizumab (95% CI 0.70, 0.88, p<0.001). Relative risk of disability progression at two years is 0.19 (95% CI 0.30, 0.12).

Table 4: Summary of key results form AFFIRM and SENTINEL studies

ITT Population	AFFIRM				SENTINEL			
Outcome	**Natalizumab (n = 627)**	**Placebo (n = 315)**	**Absolute risk reduction**	**Hazard ratio (95% CI)**	**NAT+IFN (n =589)**	**INF alone (n = 582)**	**Absolute risk reduction**	**Hazard ratio (95% CI)**
Probability of sustained disability progression (defined as an increase in EDSS sustained for 12 weeks) at two years	0.17	0.29	0.12	0.58 (0.43, 0.77)	0.23	0.29	0.06	0.76 (0.61, 0.96)
Probability of sustained disability progression (defined as an increase in EDSS sustained for 24 weeks) at two years	0.11	0.23	0.12	0.46 (0.33, 0.64)	0.15	0.18	0.03	Not given but non-significant p=0.17
Annualised relapse rate at one year	0.26	0.81	0.55	0.68 (0.59, 0.74) *	0.38	0.81	0.43	Not stated
Annualised relapse rate at two years §	0.24	0.73	0.50	0.68 (0.60, 0.74) *	0.34	0.74	0.40	0.50 (0.43, 0.59)

RES Subgroup				
Outcome	**Natalizumab (n=148)**	**Placebo (n=61)**	**Absolute risk reduction**	**Hazard ratio (95% CI)**
Probability of sustained disability progression (defined as an increase in EDSS sustained for 12 weeks) at two years	0.14	0.29	0.15	0.47 (0.24, 0.93)
Probability of sustained disability progression (defined as an increase in EDSS sustained for 24 weeks) at two years	0.10	0.26	0.16	0.36 (0.17, 0.76)
Annualised relapse rate at two years	0.28	1.46	1.17	0.81 (0.70, 0.88) *

*Relative risk reduction.

Data Syntheses

The submission does not statistically pool information about natalizumab treatment effect and this is appropriate given the different treatment and comparator regimes in the AFFIRM and SENTINEL trials, and the short term follow up in the MS201 and 231.

The submission does pool information from AFFIRM, MS201 and MS231 on safety. As the length of follow-up in these three trials is different (2 years, 12 weeks and 24 weeks) it may have been more appropriate to use rate ratios, rather than the risk ratios used in the submission. Given the shorter follow up period in MS231 and MS201, it is possible that these trials may bias the results in favour of natalizumab, as there may be less adverse effects with less exposure to the drug. Further details about how this analysis is used are given in the following section.

Indirect comparisons

In the absence of head to head trials of natalizumab with DMT comparators, the submission determines this through indirect comparison (see submission p.80-85). Comparison data is taken from the AFFIRM trial and the updated systematic reviews of INF-β and GA for RRMS. The re-calculated random effects models for the comparators are used in the indirect comparisons. Relative risks are used rather than hazard ratios for disease progression outcomes and rate ratios for relapse outcomes in the indirect comparisons. These appear to be broadly in line with the values reported from AFFIRM. Correct statistical adjustments are used to derive pairwise relative risks and we were able to replicate the results reported for disease progression. The submission recognises that the method of indirect comparison used[37] is more simplistic than a Bayesian approach, which would take into account all treatment comparisons. However, given the limited number of trials in the evidence network, this would be unlikely to have reached a substantially different result.

Results from the indirect comparison are provided (in confidence) for the following:

Comparator:

- Natalizumab vs INF-β
- Natalizumab vs GA,

Population:

- AFFIRM ITT population for natalizumab
- AFFIRM RES subgroup for natalizumab
- For both comparators, data on RRMS only is provided.

Outcomes:

- Disability progression at 2 years sustained for 12 weeks
- Disability progression at 2 years sustained for 24 weeks

For GA, progression at 2 years is defined as sustained for 12 weeks and this outcome is used in the comparison. For IFN-β both definitions of sustained progression are reported in individual trials and seem to have been combined in the meta-analysis.

- At least one relapse at 2 years

The submission argues that RES patients are less likely to respond to IFN-β because the mode of action in MS is unknown, making it is reasonable for them to assume that impact in RRMS is the same as RES. Data from the AFFIRM trial shows a slightly greater impact in the RES subgroup compared to the RRMS group. The logic of this argument is not tested and may be unsound.

For natalizumab vs GA, calculations are made both including and excluding one study from the Cochrane review of GA. This trial was excluded in the ScHARR model on which the current model is based (See Chapter 5 for greater detail of how this model is used). It is not clear why this study should be excluded. [redacted]

[redacted]
5[redacted]

Table 6[redacted]

[redacted]

Indirect comparison results for relapse rates a show significant decrease ($p<0.01$ in all cases) in the proportion of people experiencing at least one relapse at 2 years with natalizumab compared to both GA and IFN- β, and for both the ITT RRMS group and the RES subgroup (and Table 8). RR of at least one relapse at 2 years compared to IFN- β was 0.63 (95% CI 0.53, 0.77) in the RRMS group and 0.49 (95% CI 0.36, 0.66) in the RES subgroup. RR of at least one relapse at 2 years compared to GA was 0.57 (95% CI 0.45, 0.71) in the RRMS group and 0.43 (95% CI 0.31, 0.60) in the RES subgroup.

Table 5: Results of the indirect comparison of disability progression for natalizumab and IFN-beta (from table 30, p 83 of submission)

[redacted]

Table 6: Results of the indirect comparison of disability progression for natalizumab and GA (from table 31, p 84 of submission)

Table 7: Results of the indirect comparison of relapse for natalizumab and IFN-beta (from table 32, p 85 of submission)

ITT population													
Cochrane endpoints	**AFFIRM endpoints**	**Cochrane (n = 919): IFN-beta vs. placebo**				**AFFIRM (n = 942): NAT vs. placebo**				**Indirect: NAT vs. IFN-beta**			
		RR	lcl	ucl	p	RR	lcl	ucl	p	RR	lcl	ucl	p
All patients with at least one exacerbation at 2 years	All patients with at least one exacerbation at 2 years	0.81	0.74	0.89	*	0.51	0.44	0.61	*	0.63	0.53	0.77	*
RES subgroup													
Cochrane endpoints	AFFIRM endpoints	Cochrane (n = 919): IFN-beta vs. placebo				AFFIRM (n = 209): NAT vs. placebo				Indirect: NAT vs. IFN-beta			
		RR	lcl	ucl	p	RR	lcl	ucl	p	RR	lcl	ucl	p
All patients with at least one exacerbation at 2 years	RES patients with at least one exacerbation at 2 years	0.81	0.74	0.89	*	0.39	0.29	0.53	*	0.49	0.36	0.66	*

* p < 0.01; IFN-beta, interferon beta; NAT, natalizumab; RES, rapidly evolving severe subgroup; RR, risk ratio; lcl, lower confidence limit (95%); ucl, upper confidence limit (95%)

Table 8: Results of the indirect comparison of relapse for natalizumab and GA (from tables 33, p 85 of submission)

ITT population													
Cochrane endpoints	**AFFIRM endpoints**	**Cochrane (n = 251): GA vs. placebo**				**AFFIRM (n = 942): NAT vs. placebo**				**Indirect: NAT vs. GA**			
		RR	lcl	ucl	p	RR	lcl	ucl	p	RR	lcl	ucl	p
All patients with at least one exacerbation at 2 years	All patients with at least one exacerbation at 2 years	0.91	0.77	1.07	0.26	0.51	0.44	0.61	*	0.57	0.45	0.71	*
RES subgroup													
Cochrane endpoints	AFFIRM endpoints	Cochrane (n = 251): GA vs. placebo				AFFIRM (n = 209): NAT vs. placebo				Indirect: NAT vs. GA			
		RR	lcl	ucl	p	RR	lcl	ucl	p	RR	lcl	ucl	p
All patients with at least one exacerbation at 2 years	RES patients with at least one exacerbation at 2 years	0.91	0.77	1.07	0.26	0.39	0.29	0.53	*	0.43	0.31	0.6	*

* p < 0.01; GA, glatiramer acetate; NAT, natalizumab; RES, rapidly evolving severe subgroup; RR, risk ratio; lcl, lower confidence limit (95%); ucl, upper confidence limit (95%)

Adverse effects

Safety data about natalizumab made available to the FDA at the time of licensing included 1617 patients from both controlled and uncontrolled studies with a median drug exposure time of 20 months. The most frequently reported adverse effects with natalizumab were infection, hypersensitivity reaction and depression.[38]

In the AFFIRM trial data, although few differences are significant, there is a trend towards more AEs with natalizumab overall. Only fatigue and allergic reaction are shown as significantly more common with natalizumab compared to placebo (fatigue 27% vs 21%, p=0.048; allergic reaction 9% vs 4%, p=0.012) although the studies are not powered to detect such differences.

Two deaths were reported in the AFFIRM trial. Both were in the natalizumab arm but appear unrelated to treatment. One person died of malignant melanoma but had a history of this disease at the time of commencing treatment with natalizumab, the other died of alcohol intoxication.

The main serious adverse effect of natalizumab is progressive multifocal leukoencephalopathy (PML) which has been reported in three people treated with natalizumab, two of these cases were fatal. As a result, the manufacturer voluntarily suspended drug marketing in February 2005.[39] Two cases of PML occurred in the SENTINEL trial among those taking natalizumab in combination with IFN-β for MS, the third was in a trial of natalizumab for Crohn's disease. It has been suggested that the risk of PML may increase with the duration of treatment with natalizumab and the use of other immunotherapy.[40] The action of natalizumab, which suppresses the migration of T-cells and immune responses mediated by them, increases the risk of infections. In passing we note that it has been suggested that one of those suffering fatal PMS was not in fact suffering from multiple sclerosis as trial entry criteria was based on clinical criteria alone, without confirmation on MRI or examination of cerebrospinal fluid examination.[41] Data about the safety of long term natalizumab use is not yet available.

The license for natalizumab was reinstated three months later after surveillance measures were put in place.

A survey of 3417 people who had received natalizumab for a mean of 17.9 monthly doses while receiving treatment in clinical trials for multiple sclerosis, Crohn's disease or rheumatoid arthritis failed to identify any additional cases of PML on MRI or testing of cerebrospinal fluid.[42] This gives a risk of 1.0 per 1000 people treated (95% CI 0.2, 2.8 per 1000).

The EMEA scientific discussion about natalizumab outlines additional risk minimisation activities that the CHMP considered were required in relation to treating the RES group, because of infection concerns, including about PML:

- Clear cut definition of the target population, i.e. restricted use only for patients with highly active disease without reasonable alternatives.
- Requirement for established MS
- Escape rule for non-responders to avoid unnecessary exposure
- Administration only in specialised centres by experienced physicians
- Clear contraindications including a contraindication for combination with other immunomodulators
- Patient alert card
- Education programme for physicians.

Indirect comparison of adverse events

The submission used the same approach to indirect comparisons to compare safety data on natalizumab with IFN-β and GA as used for treatment benefit. However, data from the short term trials MS 201 (12 week follow up) and MS231 (6 month follow up) are also used to inform this analysis (p.89 of submission). As discussed above, it is possible that fewer adverse effects are seen over these shorter follow up periods which could bias the analysis in favour of natalizumab. Pooling using rate ratios, rather than risk ratios, may have been the more appropriate given these different follow up periods.

The indirect comparison is made more difficult by the fact that some adverse effects are classified differently in the different drug trials and so may not be comparing exactly the same outcomes.

The pooled results from the Cochrane reviews of IFN-β and GA are themselves based on trials that had different reporting strategies for AEs; some trials included in the reviews did not report AEs, AEs were measured after different periods of follow up and definitions of some AEs varied. In trials of IFN-β, different composite measures were reported variously as "flu-like syndrome" or "flu-like symptoms". There was significant heterogeneity in the meta-analysis for this outcome (chi^2=12.37, df=3, p=0.006).[6] However, a significantly greater risk of flu-like symptoms with IFN-β was found in the indirect comparison with natalizumab (RR 0.47, 95% CI 0.26, 0.82, p=0.01). One other AE, myalgia/ arthralgia was also seen in significantly more people treated with natalizumab in the indirect comparison, although this is more marginal (RR0.68, 95%CI 0.47, 0.98, p=0.04). Myalgia is one element of the "flu-like symptoms" composite outcome.

No significant differences in AEs for those treated with natalizumab compared to GA were found.

No indirect comparison is provided for allergic reaction, although significantly more people treated with natalizumab in the AFFIRM trial experienced this than those treated with placebo. The potential for treatment-related death is not considered in indirect comparison.

4.3 Summary

Evidence to support the efficacy of natalizumab as monotherapy in RRMS is based on one large, well-conducted trial, AFFIRM. This provides evidence of a reduction in disease progression and relapse rates for those treated with natalizumab compared to placebo. Evidence to support the use of natalizumab in HARRMS is based on a post-hoc analysis of a RES subgroup form this same trial. Natalizumab appears to have a similar impact on progression and relapse rate in this group with more active disease.

No randomised controlled trial has explicitly examined the impact of natalizumab as monotherapy in a SOT group. Data to support this comes from extrapolating trial data from trials of other treatment regimens or in other treatment groups, neither of which is ideal.

Thus acceptance in the evidence for SOT is based on either the assumption that the impact of natalizumab combined with IFN-β is the same as natalizumab alone (compared to IFN-β), or that a relatively treatment naive RRMS group at a later time point can be regarded as a proxy for the SOT subgroup.

As no head to head trials exist of natalizumab compared to IFN-β or GA, these have been estimated thought indirect comparison. Although the methods used to undertake these comparisons are appropriate some consider that, because of the methodological debate that exists, indirect comparisons should be regarded with caution. [redacted] Relapse rate does appear to be improved in both populations with natalizumab treatment.

[redacted] Relapse reduction appears to be significantly improved with natalizumab compared to GA.

Incidence of fatigue and allergic reaction were significantly more common with natalizumab compared to placebo. Risk of developing PML is estimated at around 1 in 1000 in a population treated for a mean of 18 months.

5 ASSESSMENT OF COST-EFFECTIVENESS

This chapter provides an assessment of the cost effectiveness analysis submitted by Biogen, plus additional analyses carried out by PenTAG. The chapter starts with a short summary of the methods used in the cost effectiveness analysis, and the baseline results presented in the submission. It then critiques the submission using standard approaches for the critical appraisal of economic evaluation and guidelines for good practice in decision-analytic modelling.[43]

5.1 Overview/summary of manufacturer's economic assessment

A systematic review of economic evaluations of natalizumab was undertaken by Biogen. This searched an acceptable core of databases (Medline, Medline in process, Embase and NHS EED). Details of the searches, including their time frame and language limits are described, and allow the searches to be reproduced. The search did not identify any economic evaluations for natalizumab.

No separate literature searches are reported for quality of life, or resource use and costs. Despite the paucity of evidence for the intervention it would be expected that a broader systematic search should be run for the population group as a whole, within the multiple sclerosis literature, to retrieve suitable model parameters.

The manufacturer submission reports cost effectiveness analyses (CEA), presenting cost per QALY estimates for natalizumab compared to:

(i) BSC (BSC reflects the placebo arm of RCTs)

(ii) IFN-β

(iii) GA

Cost per QALY estimates are presented for the RES and SOT subgroups of patients with RRMS.

5.1.1 CEA Methods

The CEA uses a decision-analytic model (Markov-process cohort model) to estimate the incremental costs and benefits associated with natalizumab treatment, versus stated comparators. The main components of this model are summarised below (see submission for detail).

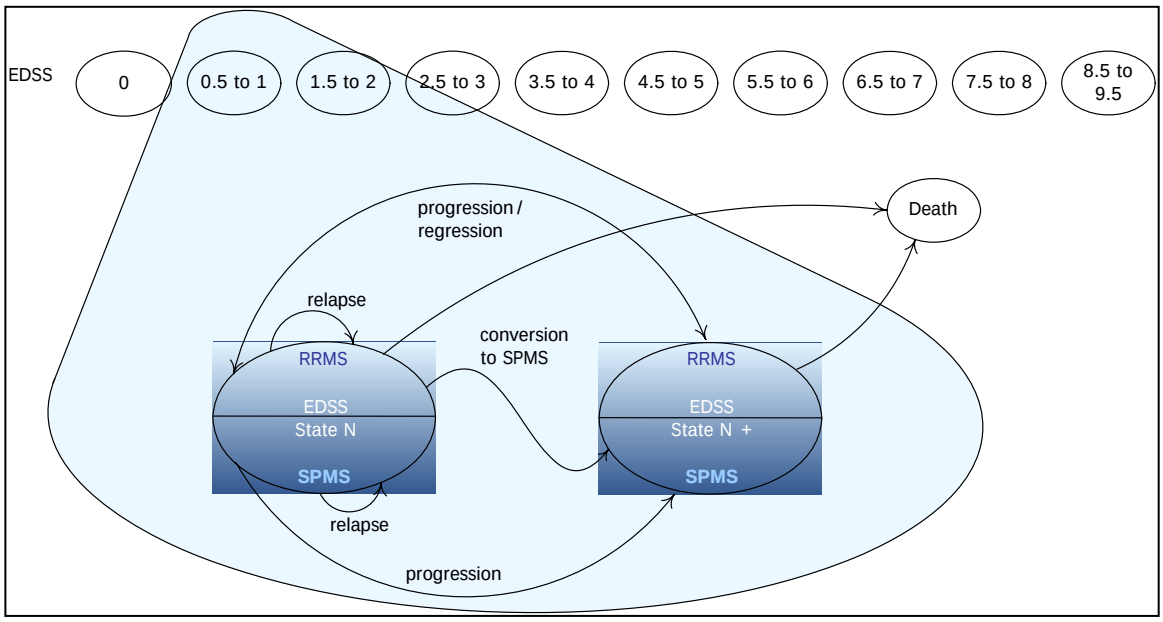
Natural history

The model predicts disability progression (as measured using the Kurtzke EDSS disability scale) and disease activity (frequency of relapse) over time. CEA is presented over a 20-year time horizon, via a series of 1-year model cycles. Disability progression is modelled using annual transitions between health states defined by EDSS 0 (normal neurological examination) to 10 (death due to MS). The relapsing and remitting nature of MS is captured in the model via a probability of having one or more relapses in each model cycle. The model considers the experiences of a cohort of patients over time (according to different treatment strategies), with the cohort beginning in RRMS and having a probability of moving to different EDSS states in RRMS, or moving to SPMS EDSS states or death (see

Figure 2).

Active treatment strategies (natalizumab, IFN-β, GA) are followed from the start of the model (all persons) up until people either progress beyond EDSS 6, transit to SPMS, or drop out of treatment for other reasons. Withdrawal rates are applied for active treatment options, with differential rates applied by treatment (natalizumab 6.4%, IFN-β and GA 5.5%) applied in each of the first 10-years of the model. When people are on natalizumab treatment the model captures adverse events, investigation of suspected PML, and testing for the presence of neutralising anti-bodies (NAB). However, no patients are withdrawn from natalizumab treatment in the model due to PML or NAB.

Figure 2: Model Structure (from Manufacturer submission, Figure 14)



Treatment effectiveness

Treatment effectiveness, the difference between natalizumab and comparator strategies, is based on data from RCTs comparing natalizumab with placebo (AFFIRM) and a systematic review of trials comparing active comparators of IFN-β and GA with placebo. An indirect comparison of data is used to estimate the relative treatment effect between active comparators. However, whilst indirect assessments are presented in the submission to NICE, these indirect assessments (submission Tables 30 & 31) are not used in the CEA.

The AFFIRM trial reports a reduced risk of sustained progression of disability for people treated with natalizumab compared to placebo (model uses: HR of 0.36 for RES, HR of 0.46 for SOT), over 2 years. The submission reports findings from meta-analysis reporting a reduced risk of sustained progression of disability for those treated with IFN-β or GA compared to placebo (RR of 0.70 for IFN-β , RR of 0.88 for GA). The AFFIRM RCT reports a reduced rate of relapse for those treated with natalizimab compared to placebo (model uses: RR of 0.192 for RES, RR of 0.321 for SOT).

The submission calculates an annualised relapse rate for IFN-β compared to placebo, and GA compared to placebo, from published Cochrane reviews (model uses: RR of 0.667 and 0.745 for IFN-β RES and SOT groups respectively; RR of 0.710 and 0.782 for GA RES and SOT groups respectively). These data show considerable differences between natalizumab and IFN-β and GA. These are the data that are used in the model to modify the progression of MS in the cohort analysis. These data are applied to a common model of expected disability progression and relapse activity for the control group (BSC).

Resources and costs

The model estimates additional treatment costs (drug costs, administration costs, costs of treating of adverse events) for each active comparator. For other longer-term costs the model uses estimates of the costs for EDSS health states. State-occupancy data from the model are used to capture the longer term cost consequences of MS (throughout the time horizon of the model). EDSS health state costs are based on data from the UK MS Survey 2005 (submission cites related study; Tyas and colleagues, in press).

Health Related Quality of Life

The analysis estimates differences in QALYs gained between treatment strategies. It applies health state utilities for each EDSS state in the model, and captures differences in the cohort over time based on health state occupancy. The health state values are based on EQ-5D data collected in the UK MS Survey 2005 (submission cites related study; Orme and colleagues, in press).

The analysis also uses a decrement in QALY values for each relapse experienced in the cohort analysis (ranging from -0.138 in EDSS 0 to -0.024 in EDSS state 5+). A relapse is assumed to last 46 days and the health state value is adjusted over this period to reflect the disutility associated with relapse.

The impact of treatment options on HRQL is reflected in the analysis using a standard disutility estimate per treated patient in each model cycle.

The model uses a disutility of -0.013, -0.047 and -0.007 per year for GA, IFN-β and natalizumab respectively. This measure of disutility is used to capture the disutility associated with treatment options and the different modes of administration (i.e. lower frequency of administration for natalizumab). A very small utility decrement (0.00039 to 0.000146 QALYs per year per person treated) is used in the model to cover adverse events whilst on natalizumab.

An estimate of the impact of treatment on the utility of caregivers is also used in the analysis. Caregiver disutility is derived from UK MS Survey (2005) data on time spent by caregivers, plus data from other sources. Estimates of caregiver disutility vary by EDSS state (disability) over a range of 0 (EDSS 0-2) to 0.14 (EDSS 9).

Discount rates (costs and benefits)

A discount rate of 3.5% is used for future costs and life years/QALYs.

Sensitivity analysis

Extensive one-way sensitivity analysis is reported, plus multi-way sensitivity analysis and probabilistic sensitivity analysis (PSA).

Model validation

The submission reports consideration of model validation, including tests for internal consistency, and consideration of the model in the context of model inputs (transit probabilities) and other analyses of DMTs for MS.

5.1.2 Results of cost-effectiveness analysis

CEA results are presented as incremental cost per QALY gained for the RES and SOT treatment subgroups (Table 9). Results presented here were supplied by the manufacturer in an updated analysis (Addendum to submission dated 22/12/06), following our identification of miscalculations in the original spreadsheet model.

The base case analysis presents an estimated cost per QALY of between £32,000 and £44,600 for natalizumab versus active comparators and BSC in the RES subgroup. Cost per QALY estimates are higher for the SOT subgroup with estimates ranging from £43,300 to £56,100 for natalizumab versus active comparators and BSC.

Table 9: Cost effectiveness results presented in industry analyses, for base case analyses over 20-year time horizon

Analysis	**Cost per patient**	**QALYs per patient**	**NAT versus comparator Incr. cost**	**NAT versus comparator Incr. QALYs**	**Estimated cost per discounted QALY: NAT vs. comparator**
RES:					
NAT	£162,000	7.51	-	-	-
BSC	£84,700	5.78	£77,300	1.73	£44,600
IFN-β	£122,300	6.27	£39,800	1.24	£32,000
GA	£110,000	6.01	£52,100	1.50	£34,600
SOT:					
NAT	£159,500	7.58	-	-	-
BSC	£79,200	6.15	£80,300	1.43	£56,100
IFN-β	£119,200	6.65	£40,300	0.93	£43,400
GA	£106,200	6.38	£53,300	1.20	£44,300

Table 10 reports selected one-way sensitivity analysis (see full details in manufacturer Addendum, Table 85). A wide range of parameters have an impact on the cost per QALY estimates. The time horizon shows the greatest impact, with analysis using disability progression data from the London Ontario data also having considerable impact on cost effectiveness. Other parameters highlighted by the manufacturer as having the greatest impact on the ICER are effectiveness data (for natalizumab and comparators), decision-making perspective, and baseline characteristics. On decision-making perspective, the large societal costs included in the sensitivity analysis lead to a very much reduced ICER when a societal decision making perspective is adopted. For example, when considering the RES subgroup, natalizumab versus GA, in the base case scenario the mean health state costs per patient increase 4-fold (from £59,600 to £239,000) in the natalizumab cohort, and by almost the same magnitude in the GA and BSC cohort (to £293,000 and £304,000).

Table 10: Sensitivity analysis presented in the manufacturer submission (Selected from Addendum Table 85), Cost per QALY in £'000s (Nat vs comparator)

Base case	Scenario	RES NAT vs. IFN-β (£K) 32.0	SOT NAT vs. IFN-β (£K) 43.4
Time horizon – 10-years	10.1	55.2	69.8
Time horizon – 30-years	10.2	24.6	34.2
Progression data from London Ontario data	4.8	42.3	55.3
Mean age at baseline = baseline +20-years	1.3	36.6	48.3
NAT effectiveness data: disease progression data from 12=weeks endpoint	5.1	42.8	68.3
NAT effectiveness data (progression), lower SE used	5.2/5.3	52.6*	57.3*
NAT effectiveness data (progression), upper SE used	5.2/5.3	25.6**	35.9**
Utility for EDSS health states – at upper 95% limit	7.1a	37.4	52.9
Resource use and drug costs for IFN – upper 95% limit	8.1a	39.9	59
Resource use and drug costs for IFN/GA – lower 95% limit	8.1b	19.2	29.1
Perspective – societal costs perspective	9.1	11.4	23.6
Perspective – Govt costs perspective	9.2	25.1	37.2

* Hazard ratios used were 0.25 for RES, 0.39 for SOT

** Hazard ratios used were 0.39 for RES, 0.54 for SOT

The manufacturer submission presents sensitivity analysis whereby effectiveness data from the RES subgroup are applied to the SOT subgroup, with subsequent adjustment to age (assuming the SOT group represent a subgroup of RES at a later point in time). The results from this multi-way sensitivity analysis, with RES disability progression rates and efficacy applied to the SOT group, show a cost per QALY of £32,000 for natalizumab compared to IFN-β, £35,300 for natalizumab compared to GA, and £44,600 for natalizumab compared to BSC.

The CEA presents results for PSA. For the RES subgroup, baseline CEA scenarios, it reports that where there is a willingness to pay of £30,000 per QALY natalizumab compared to IFN-β will be cost effective in 42% of cases, and in 32% of cases when compared to GA.

For the SOT subgroup, baseline CEA scenarios, the PSA reports that where there is a willingness to pay of £30,000 per QALY natalizumab compared to IFN-β will be cost effective in 18% of cases, and in 15% of cases when compared to GA (CEACs, and CEA planes are presented in the analysis).

5.2 Critical appraisal of manufacturer's economic evaluation

5.2.1 Critical appraisal of economic evaluation methods

We considered the methods applied in the economic evaluation against the critical appraisal questions listed in Table 11, drawn from widely used tools for assessment of economic evaluations (e.g. Drummond and colleagues 1997[43]).

Table 11: Critical appraisal of submitted economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	✓ - Yes	None
Is there a clear description of alternatives (i.e. who did what to whom, where, and how often)?	✓ - Yes	Natalizumab vs placebo IFN-β vs placebo GA vs placebo
Has the correct patient group / population of interest been clearly stated?	? – Yes/No	Patient groups (RES & SOT) are based on the licensed indication. See section 2.3 for discussion.
Is the correct comparator used?	✓ - Yes	MTX in the scope but not submission, although justification provided.
Is the study type reasonable?	✓ - Yes	CEA model used, CUA results presented.
Is the perspective of the analysis clearly stated?	✓ - Yes	Analysis presented from different perspectives; including UK NHS & PSS
Is the perspective employed appropriate?	✓ - Yes	Submission presents UK NHS and PSS perspective (consistent with NICE reference case).
Is effectiveness of the intervention established?	?	The CEA is based on clinical effectiveness data from the AFFIRM trial (NAT). See discussion in section 3.2.1. Detail is presented on the RCT, but interpretation of trial findings, and consideration of methods employed in the CEA are required by NICE.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	✓ - Yes	CEA uses 20-year time horizon, this is not a lifetime horizon, but it is able to capture longer term disease progression, and is consistent with previous models developed for NICE in appraisal of MS treatment.
Are the costs and consequences consistent with the perspective employed? *	? – Yes/No	For NHS & PSS, costs are consistent with approach taken (see discussion). Care giver disutility included in base case analysis for NHS & PSS perspective, this may not be appropriate for NICE reference case.
Is differential timing considered?	✓ - Yes	None
Is incremental analysis performed?	✓ - Yes	None
Is sensitivity analysis undertaken and presented clearly?	✓ - Yes	None

* More on data inputs for costs and consequences in the review of modelling methods below

5.2.2 NICE reference case requirements

Table 12 compares the manufacturer's submission to the requirements of the NICE reference case. There is some overlap with the items presented above on the general approach for the economic evaluation.

Table 12: Assessment of submission against NICE reference case requirements

NICE reference case requirements (see detail in NICE report):	Included in Submission
Decision problem: As per the scope developed by NICE	✓(?)*
Comparator: Alternative therapies routinely used in the UK NHS	✓
Perspective on costs: NHS and PSS	✓**
Perspective on outcomes: All health effects on individuals	✓
Type of economic evaluation: Cost effectiveness analysis	✓
Synthesis of evidence on outcomes: Based on a systematic review	X (?)***
Measure of health benefits: QALYs	✓
Description of health states for QALY calculations: Use of a standardised and validated generic instrument	✓
Method of preference elicitation for health state values: Choice based method (e.g. TTO, SG, not rating scale)	✓
Source of preference data: Representative sample of the public	✓****
Discount rate: 3.5% pa for costs and health effects	✓

* The NICE scope states MTX as a comparator, but industry submission does not include in their analysis (a rationale for this decision is provided in the industry submission).

** Other perspectives also explored

*** RCT data for NAT, meta-analysis for IFN & GA. See section 3.2 for rationale provided.

**** This is the case for utility estimates for health state utilities, other non representative sample use for disutility of treatment

5.3 Critical appraisal of cost-effectiveness modelling methods

A general critical review of modelling methods has been undertaken. The review has used the framework for good practice in modelling presented by Philips and colleagues (2004) as a guide,[44] addressing issues of model structure, structural assumptions, data inputs, consistency, and assessment of uncertainty.

5.3.1 Modelling approach and structure

The model structure is based around the use of the EDSS to capture disability associated with MS. The model structure also captures the relapsing nature of MS using a relapse rate applied in each cycle of the model. The model is structured around RRMS and SPMS. The model structure is based around earlier models developed for use by NICE in the appraisal of DMTs for MS (referred to as the ScHARR model).

Whilst there is some dissatisfaction with the use of EDSS to capture disability in MS[45] (see Section 4.1.6), given the nature of the evidence base in MS it seems reasonable to use the EDSS in this instance. The limitations of the EDSS to accurately reflect MS disability and to reflect the broader HRQL associated with MS should be considered in interpreting the model findings. However, the approach to modelling MS for the purposes of cost-effectiveness appears to be appropriate given the decision problem and the data available.

The model is presented in a spreadsheet format (Excel). A 20-year time horizon, and 1-year cycle length are used (a form of half-cycle correction is used in the analysis). The model uses 21 health states. Health states are based on the EDSS state descriptions, using 10 health states each for RRMS and SPMS, and a further health state is used for death. The EDSS comprises 20 defined states. However, in this model EDSS states are rounded up to band states together (e.g. health state for EDSS state 4 includes EDSS 3.5 and EDSS 4.0), giving 10 health states for the range of MS disability.

Whilst the manufacturer submission states that the model structure is based on the previous ScHARR model there are a number of important differences. The models are both based on transition matrices between EDSS health states, for RRMS and SPMS. However, the ScHARR model uses the full continuum of 20 EDSS health states (0 to 10, by increments of 0.5). This may impact on the assessment of the cumulative probability of sustained progression of disability, using the trial endpoint defined in AFFIRM, given that no moves are possible other than across a full 1.0 point shift in the EDSS scale.

The manufacturer model uses data from the AFFIRM trial to populate transit matrices for EDSS states 0 to 6 for RRMS, whereas the ScHARR model used the London Ontario dataset. It is important to note that in the natalizumab model these data are from relatively small samples: the RES placebo group contains 61 people and the SOT placebo group 315. AFFIRM data was collected over a relatively short 2-year time period (London Ontario data are over 25 years). The data used dictate that it is possible for people to move from their current EDSS health state to an improved (reduced disability) EDSS health state. The probabilities associated with these transits to improved EDSS health states are very high in a number of cases (see discussion below in section 'Data'). In the ScHARR model transitions to improved disability EDSS states were not possible.

The 1-year cycle length used in the model is appropriate. The half-cycle correction applied in the model weights the annual health state costs and utilities to reflect the occupancy of the health states at mid-cycle. The time horizon of the analysis is 20-years, and whilst this is not a life-time horizon it would seem a reasonable period for analyses. The time horizon is consistent with the previous analytical approach employed by NICE (other time horizons are employed in sensitivity analysis).

The manufacturer submission summarises a number of other differences between the model presented and the ScHARR model (see p154 of submission). We note that a number of these issues are of no interest and are only relevant for comparing results between models, such as differential discount rates. However, we have run additional sensitivity analysis against the different approach to withdrawal in the ScHARR model, and the different approach to disutility of treatment (see below, PenTAG sensitivity analysis). We find that using alternative assumptions for withdrawal and disutility of treatment has a minimal impact on cost effectiveness (one-way sensitivity analysis).

5.3.2 Structural assumptions

The model is structured around the use of a common baseline model of MS for BSC through which each active treatment can be compared with BSC.

Where active comparators are compared against each other (e.g. natalizumab vs. IFN-β) each treatment is compared to the BSC cohort, and differences between the two active treatment versus BSC comparisons are used to estimate the incremental cost per QALY between the active treatment options being compared.

Crucial to the CEA is the use of transition probabilities to predict disability progression between EDSS health states in the BSC cohort. This baseline disability progression is modified (in each 1-year cycle of the model when on treatment) through the use of the relative risk or hazard ratios of treatment compared to placebo. For example, where the cohort has a 50% probability of progressing to a higher (worse) EDSS state in a particular 1-year cycle, natalizumab treatment in the RES and SOT groups would reduce this probability of disability progression to 18% and 23% respectively. As well as reducing the risk of progression, a greater number of people are left in their current health state, thereby increasing the number (compared to BSC) subject to a probability of an improved EDSS state in the next 1-year cycle.

The use of a baseline BSC model of disability progression places much importance on its development and validity. Whilst the manufacturer submission has considered model validity, we have undertaken analysis which indicates that the model is predicting a different rate of disability progression to that reported in the AFFIRM trial. Our analysis (see section 5.3.4) also indicates that the model leads to a much greater treatment effect than that reported in the AFFIRM trial, through the use of relative risks/hazard ratios to modify the underlying model of disability progression. We believe that the use of data from the AFFIRM trial to derive transition matrices (for RRMS EDSS states 0-6) may lead to some asymmetry between the model predictions and the outcomes reported in AFFIRM. Patient level data on 3-monthly assessments of EDSS is used to model transition probabilities. This data may not reflect the primary trial outcome measure of the cumulative probability of progression of disability sustained for 24 weeks at two years. The fact that the model does not use the full continuum of EDSS health states, banding EDSS states together may also be an important consideration. Our analysis, presented in more detail below (under consistency), is simple and illustrative. However, the issue is important because if the model predicts a greater (more rapid) rate of disability progression, the predicted benefits from treatment aimed at reducing disability progression may be overstated.

As discussed above, unlike the ScHARR model, one structural assumption in the model is the use of transition probabilities that permit backward transitions to improved EDSS health states, when people are in RRMS health states of EDSS 0-6. This is based on EDSS data from the AFFIRM trial. However the manufacturer submission does not discuss the rationale for the difference in modelling approach in any detail.

A further structural assumption is the adjustment of the probability of people moving from RRMS to SPMS at each 1-year cycle. The model applies a relative risk reduction parameter to modify the risk of transition from RRMS to SPMS in each cycle. An assumption is made to use 50% of the risk reduction seen in the RRMS strata of the model. This assumption is based on a fitting of the model and trial data (discussed in Section 6.2.12.3 of the manufacturer submission). However, the rationale for this assumption is not clearly stated.

We have undertaken additional analysis which strongly suggests that the disability progression element of the model accounts for the vast majority of the treatment impact (and subsequent cost-effectiveness estimates), with the relapse rate having a small impact on cost-effectiveness (see Appendix 6, page 126). We have therefore devoted more time to considering the disability progression element of the model (and CEA). However, the estimation of baseline relapse rates and the adjustment of the baseline disease activity using relative risks data for each of the active treatments is also of some concern (e.g. use of data from multiple sources to estimate relapse rate, use of data from UK MS Survey 2005), although these issues have not been explored in any detail.

5.3.3 Data inputs

Patient group

The analysis uses a MS patient cohort defined from AFFIRM and other sources. Patients have a mean age of 36 years, a mean time from diagnosis of 5-years, the cohort predominately contains women (with a 3:1 ratio). The distribution of the cohort at the start of the model by EDSS state is taken from the placebo group of the AFFIRM trial – all with RRMS across EDSS 0 to 6, with the majority of the cohort in EDSS 2 (37%), 3 (23%) and 4 (16%).

Clinical effectiveness

The treatment effect is modelled over time. As described above, the data from the AFFIRM trial on the relative reduction in the risk of progression of disability is used to modify the predicted baseline BSC progression of disability associated with MS. The data from AFFIRM on relapse rates for natalizumab is also applied. Data on the clinical effectiveness of active comparators (IFN-β, GA) are taken from published studies and this is reviewed in Section 3.2. These clinical effectiveness data are used to model differences across treatment strategies through the delay of disability progression, the reduction in relapse rates, the differences in withdrawal rate across active treatments, and the differences in adverse event profiles (cost and QALY differences). Table 13 presents the data used in the model to modify the risk of disability progression and relapse (base case analysis). For IFN-β and GA the same relative risks are applied for disability progression in RES and SOT analyses.

Table 13: Clinical effectiveness data applied in the cost effectiveness analysis (base case scenarios), relative risks/hazard ratios

	RES		SOT	
Treatment	Disability Progression* RR/HR (95% CI)	Relapse** RR/HR (95% CI)	Disability Progression* RR/HR (95% CI)	Relapse** RR/HR (95% CI)
NAT	0.36 (0.17, 0.76)	0.192[+] (0.12,0.30)	0.46 (0.33, 0.64)	0.321 (0.26, 0.40)
IFN-β	0.70 (0.55, 0.88)	0.667 (0.74, 0.89)[#]	0.70 (0.55, 0.88)	0.745 (0.74,0.89)[#]
GA	0.88 (0.56, 1.38)	0.710 (0.63, 1.12)[#]	0.88 (0.56, 1.38)	0.782 (0.63, 1.12)[#]

* Nat = probability of sustained disability progression (defined as an increase in EDSS sustained for 24 weeks), at 2-years (compared to placebo); IFN-β/GA = data from meta-analysis, disability progression at 2-years (compared to placebo).

** relative relapse rates – CIs not supplied for relative relapse rates in submission table 69

[+] a rate of 0.194 is stated in the submission (Table 70)

these CIs relate to the relative risk of relapse rather than the relative relapse rate quoted, but are used in the model to sample values for the PSA.

Data used to specify the natural history of MS for BSC are drawn from a number of sources. The probability of movements across EDSS states for RRMS are derived from data from the AFFIRM trial placebo group, supplemented with data from the London Ontario dataset (a long term observational study of people with untreated MS).

The data used to derive probabilities for moving from RRMS to SPMS are from the London Ontario data set. Probabilities for disease progression in SPMS are taken from the London Ontario dataset.

Relapse rates depend on EDSS state, and are typically higher at lower EDSS. Relapse rates are derived using data from multiple sources (from the UK MS Survey 2005, supplemented with some data from AFFIRM, and adjusted using relative incidence data from a study published by Patzold & Pocklington 1982[46]). In each 1-year cycle of the model patients are at risk of death. This is based on the use of standard mortality rates (from population life-tables, by age), which are adjusted by level of MS disability (mild, moderate or severe, by EDSS groups) on the basis of a published epidemiological study (Pokorski 1997[47]).

Disability progression from AFFIRM (natalizumab) and other published studies (IFN-β, GA) is modified and extrapolated beyond RCT time horizons. The model applies a constant treatment effect (RR/HR) in each cycle of the model when patients are on active treatment. Our analysis of state-occupancy in the model over time, (see Appendix 7.6), indicates that approx 34%, 25% and 20% of the natalizumab treatment cohort (in the SOT subgroup), are still receiving active treatment and so the constant treatment benefit, at 10-years, 15-years and 20-years respectively (there is a similar profile for RES). The ScHARR model also applied a constant relative risk reduction over a 20-year time horizon. However, we do not know the number of patients still receiving treatment and its effect in the later cycles of the ScHARR model. In the present manufacturer model the IFN-β and GA SOT cohort analyses have approximately 13% and 9% still on treatment at 20-years.

As well as concerns over the extrapolation of treatment effect in a 'time-independent' manner, we also have concerns over the magnitude of the treatment effect predicted in the model compared to that reported in the AFFIRM trial. Our analysis indicates that the use of the effectiveness data to modify the predicted natural history for BSC, predicts a much greater treatment effect than that seen in the AFFIRM trial (see Section 5.3.4 below).

Whilst our checking for accuracy is supportive of the mathematical methods used to estimate transit probabilities for RRMS health states 0-6 (from the AFFIRM data presented in the submission), we are unable to consider the data used from the London Ontario dataset, as it is not in the public domain. Of note is the use of the data from the London Ontario dataset to predict disability progression in a large part of the model (RRMS states 7-9, RRMS to SPMS, SPMS health states). Whilst the London Ontario dataset has been used widely in the analysis of the natural history of MS, it is taken from a long term observational dataset largely comprising untreated MS patients, and it may not reflect the HARRMS patient group relevant for the current appraisal (CEA). The manufacturer submission states elsewhere that the London Ontario data are not appropriate for HARRMS (submission p.94). Transition probabilities are adjusted to accommodate data from a number of sources (RRMS health states). We note that this reduces the importance of the AFFIRM data through rescaling to accommodate the London Ontario data on transits from RRMS to SPMS from each of the EDSS health states.

Above, we have discussed the fact that the current model allows people to transit backwards to improved EDSS health states. We have concerns that in some instances these 'improvement' transits appear high. [redacted] Given the presentation of the model we have not been able to re-run analysis where such 'improving' transits are not possible in the RRMS health states (0 to 6). The ScHARR model assumes disability progression is 'uni-directional' with no backward (improving) disability movement possible, given that this is "the current understanding of the disease".[48]

A further structural assumption is the use of the 'current' EDSS health state to capture a residual probability of disability progression when transition matrices are modified (to reflect clinical effectiveness) in the cohorts on active treatments. For example where a treatment reduces the risk of transition to a higher (worse) EDSS health state, it subsequently increases the chance (residual from disability progression risk reduction) of staying in the 'current' EDSS health state.

This leads to a greater proportion of the treatment cohort being in a position to improve in terms of EDSS health states (improving transits) in the next model cycle.

Given our analysis on rate of disability progression for the natural history model (BSC cohort), and the impact of treatment effect, we speculate that a number of the issues discussed above (in the context of clinical effectiveness data, and its use) may contribute to the differences noted between the modelled disability progression and the treatment effect reported in the AFFIRM trial.

We note that the BSC relapse rates used in the model are estimated from self-report data from respondents in the UK MS Survey 2005. The data from the UK MS Survey (see submission Tables 46 & 47) are used to estimate an annual relapse rate, by EDSS, for both RRMS and SPMS. Data from the MS Survey are sparse for EDSS states 7 to 9 and these states were combined in the analysis. Data from a previously published study[46] are used to used estimate a relative rate per patient per EDSS state. However, we note that the study used is based on relatively small numbers of patients (n=102), where data were collected between 1976 to 1980, with a mean follow-up period of 1,279 days. In this published study 54 of the 102 sample had had MS for 3-years or less. Data from AFFIRM are used to adjust relapse rates by MS type, with the RES group predicted to have a higher relapse rate (multiplier of 1.98 applied to RES group). In the AFFIRM trial the mean annual relapse rate at 2-years was 1.46 for the RES placebo group, whilst the model uses rates ranging from 0.972 to 1.448 (see submission Table 48). For the SOT subgroup the model uses relapse rates ranging from 0.490 to 0.729, compared to an annual relapse rate in the AFFIRM ITT group (SOT proxy) of 0.73. Whilst there may be uncertainty in the estimation of relapse rates, relapse data appear to have only a small impact on cost effectiveness results.

Clinical effectiveness data from RCTs has been used to inform the rate of withdrawal (drop out) in the analysis, with 6.4% of patients per year assumed to withdraw from active natalizumab treatment (data from AFFIRM) and 5.5% assumed to withdraw from IFN-β or GA treatment (from Herndon 2005).[49] These withdrawal rates apply for the first 10 years of the analysis, after which there is no annual withdrawal from treatment strategies. Patients that progress to SPMS are also classed as drop-outs, and are no longer treated. Treatment is withdrawn from all patients in EDSS > 6.

The ScHARR model assumed a rate of 10% withdrawal from active therapy in years 1 and 2, with a 3% rate applied thereafter (see PenTAG sensitivity analysis).

Patient outcomes

The model captures treatment differences (by strategy) in terms of health-related quality-of-life (HRQL) using QALYs. As above, QALY differences are estimated from four areas in the model; (1) disutility associated with treatment (and adverse events), (2) disutility associated with each relapse, (3) caregiver disutility, and (4) health state utilities associated with each EDSS health state. This last of these areas is by far the most important of the QALY calculations in the model (i.e. has greatest impact on cost effectiveness estimates). The model attaches a health state utility to EDSS states and, through state-occupancy over time, estimates the QALYs associated with each treatment strategy. The health state utilities for each EDSS state are based on data reported in the UK MS Survey 2005.

The comparative data from the previous ScHARR MS model are not available in the public domain, and we are unable to compare the data used here with that used in the ScHARR model. However, the report of the ScHARR model by Tappenden and colleagues (2001)[48] does present estimates of utility by EDSS state, and cost by EDSS state. These estimates were derived from data provided by the Multiple Sclerosis Research Trust, and were used to estimate cost-effectiveness data presented in an addendum to the ScHARR report. These data are presented for information in

Table 14, although we do not have information on the estimation methods used for the data.

Table 14: Manufacturer model estimates of utility (submission Table 51) for different EDSS states, UK MS Survey 2005 (with supplementary data from ScHARR report added by PenTAG)

EDSS State	SUBMISSION RRMS	SUBMISSION SPMS	ScHARR Report RRMS	ScHARR Report SPMS
0	0.91	0.87	0.959	0.874
0.5 to 1	0.84	0.80	(EDSS 1) 0.688	(EDSS 1) 0.603
1.5 to 2	0.74	0.70	0.688	0.603
2.5 to 3	0.61	0.57	0.645	0.560
3.5 to 4	0.65	0.61	0.61)	0.527
4.5 to 5	0.56	0.51	0.581	0.496
5.5 to 6	0.49	0.45	0.538	0.453
6.5 to 7	0.44	0.39	0.477 to 0.343	0.392 to 0.258
7.5 to 8	-0.01	-0.05	0.343 to 0.232	0.258 to 0.147
8.5 to 9.5	-0.15	-0.19	0.232 to -0.135	0.147 to -0.220

Note: source of ScHARR reported data is MS Research Trust, methods unknown.

The UK MS Survey is a cross-sectional postal survey to assess the resource requirements of people with MS and utility associated with disease. The survey was funded by Biogen. It is a UK specific study, but draws heavily on a European study published by Kobelt and colleagues (2006).[50] Whilst the UK MS Survey 2005 provides additional useful information to the sparse literature on MS, we have concerns over the use of the data from the UK MS Survey due to the potential for selection bias, and the issue of generalisability of data from the study to the broader MS treatment population and specifically to the CEA for the RES and SOT subgroups. Although the sample size appears large (n=2048), the response rate for the MS Survey, at only 16%, is low. The response rate may introduce selection bias, and (as Orme and colleagues have noted) a possible 'volunteer effect'. The diagnosis of MS type and relapse status was made by the respondent and not based on a confirmed clinical diagnosis. The questionnaire for the survey was distributed by the MS Trust to those MS patients known to the Trust and present on its database. The mean age was 51.4 years, and mean age at diagnosis was 38.8 years, with 53% of respondents between 50-69 years. In terms of disability status 60% of respondents were between EDSS 4 and EDSS 6.5, and 19.1% were between EDSS 7 and 9.5. Very few responses were available for EDSS states 0, 3 and 9 (Orme and colleagues).[51] Almost 30% had had a relapse during the 3-months prior to the survey.

Disease type was RRMS in 35.5%, SPMS in 37.2% and PPMS in 27.3%. No data is presented on the characteristics of the non-responders. In the survey disease severity was measured using the Adapted Patient Determined Disease Steps (APDDS), a self-rated scale, and is reported by EDSS strata (a clinician rated scale) to aid comparison with other studies (Orme and colleagues, in press).[51]

As part of the UK MS Survey EQ-5D data were collected. The EQ-5D health state classification responses from the survey have been used together with population tariff values (Dolan & Gudex 1995)[52] to estimate EQ-5D single index health state values, and thereafter subsequent regression analysis has been undertaken to estimate the health state value by EDSS health state.

The data presented by Orme and colleagues (in press)[51] are not directly comparable to those in the Table above, as they are in different bandings of EDSS states (e.g. EDSS 5-5.5 rather than 5.5-6.0 above), and some amendments have been made to the modelling results presented in the manufacturers submission.

Orme and colleagues highlight that whilst much of the data from the UK MS Survey do compare favourably with other studies (Parkin and colleagues,[53] Forbes and colleagues[54]), there are a number of health states which have marked differences, with the UK MS Survey data suggesting considerably lower utility values for EDSS states 3 and 5. Whilst the EQ-5D is consistent with the health state valuation techniques suggested in the NICE reference case, there is evidence that the EQ-5D may not have good coverage of the quality of life domains relevant to people with MS (Gruenewald and colleagues 2004,[55] Mitchell and colleagues 2005[56]).

The CEA model uses disutility associated with relapse rates, administration and adverse event profiles for each active treatment, and caregiver disutility. These model inputs have a very limited impact on the cost-effectiveness results (see sensitivity analysis of the submission, PenTAG sensitivity analysis), and we have not explored these parameter inputs in detail. We note that the utility differences associated with administration favour the use of natalizumab (with a relatively big difference between natalizumab and other treatments), and that there are a number of concerns with the treatment disutility estimates used.

Firstly, whilst presented as capturing disutility of treatment *and* adverse events, disutility for adverse events is captured in the model in a different area, with a very minimal impact per patient on natalizumab treatment (0.00039 to 0.000146 per person per year). Disutility for adverse events related to IFN-β and GA is not captured in the model. The estimates for treatment disutility from Prosser (2003)[57] are based on treatment states only. We do not find data to support the assumption that a monthly administration offers greater utility advantages over a more frequent dosing regime. Such differences should be explored further and it is not straight-forward to assume people would be prepared to take a risk of death (SG technique), or forgo future life-expectancy (TTO technique), for ease of drug administration. In our opinion the study used to inform the treatment disutilities applied in the model, Prosser 2003, does not present data robust enough to support a differential utility estimate by mode/frequency of administration. The sample sizes for each of the treatment options discussed in the submission are small (n=18 to n=20), and the lower value for option B (used to inform a higher disutility for IFN-β) could be due to non-significant variation in the small samples. We also note that the estimation of caregiver disutility in the manufacturer submission is based on sparse data, and a number of assumptions and further research is required in this area.

Mortality

The analysis uses standardised (population) mortality statistics (SMRs), all cause mortality statistics, and adjusts them for MS specific mortality using empirical data from a Canadian study. The adjustment factors used are split by 3 categories of disability severity; mild (EDSS 0-3) by a factor of 1.6, moderate (EDSS 4-6) by a factor of 1.84, and severe (EDSS 7-9) by a factor of 4.44. These data are from a Canadian study addressing the relationship between MS severity and life-expectancy in 2,348 patients followed in MS clinics during 1972-1985. In this data set there were 115 deaths observed compared to an expected 58 deaths (Sadovnick 1992, cited by Pokorski 1997[47]). Whilst the assumptions in the model seem reasonable to us, it is important to note that the small number of deaths in the dataset lead to uncertainty in the parameters used. Additional sensitivity analysis undertaken here indicates that the cost-effectiveness of treatments is not sensitive to a doubling/halving of the SMRs used (as they are small population mortality statistics), or adjustment to the adjustment factors themselves (in isolation).

However, where a large estimate of standardised annual mortality is employed (e.g. 5% mortality per year) the cost per QALY increases by over £8,000.

Our analysis of state-occupancy (Appendix 7.6) indicates that mortality in the cohort model is low, and similar for all active treatment options. The analysis reports approximately 10% of the cohort are dead at 20-years. Our analysis indicates that the patient flow through the model predicts a small mortality benefit for all treatments compared to BSC; with natalizumab showing a greater mortality benefit than IFN-β or GA.

Resource use and costs

Treatment Cost

Natalizumab treatment cost is associated with drug costs (£1,130 per vial, per 4-weeks), and an administration cost of £1,062 per year. The administration cost is based on a half day visit to a neurology clinic/ward. This seems a reasonable estimate, but may be an underestimate when considering the divisibilities of clinic visits/times i.e. cost for a day admission to a neurology clinic may be the same regardless of the fact that the infusion of natalizumab may be less than a full day. Treatment cost for IFN-β or GA is based on drug costs from the Department of Health risk sharing scheme, and data from the UK MS survey 2005 (reporting drug cost differences by EDSS state). See Table 8 of the manufacturer submission shown below.

Health state (EDSS) costs – *longer term cost consequences*

The model estimates longer term costs for cohorts using a mean cost per EDSS health state and state occupancy data within the model. The health state costs (see Table 15) are estimated based on findings from the UK MS Survey 2005. Cost analysis using data from this survey (using a 'seemingly unrelated regression' which considers the covariance structure between elements of costing) estimates the quantity of each resource per person, by MS type, EDSS state, sex, presence of relapse, then multiplied resource used data by unit costs (Submission, Table 49 [some errors in the presentation of the table]). The regression analysis estimates a reference case cost, and differences in cost when altering the reference case characteristics (e.g. EDSS state). The reference case covers age=0, female, RRMS, EDSS 0, no relapse, no DMT.

Costs by EDSS health state are used in the model presented, cost estimates used are independent of gender. Health state costs for SPMS are only marginally greater than RRMS costs, at £56 per patient per year.

The UK MS Survey 2005 has been discussed above (under patient outcomes, utility data), and whilst we accept that the data from the survey adds to a sparse literature on the costs associated with MS, we have concerns over the use of the resource use data from the UK MS Survey due to the potentially unrepresentative nature of the data for MS patients in general, but specifically for HAARMs. We have some concern over the methods used in the MS Survey to collect data and estimate annualised resource use (i.e. collecting data over 1-month or 3-month periods and extrapolating to 1-year period). For example, Tyas and colleagues (in press) report the UK sample with a mean of 45.4 consultations (medical and paramedical visits), plus a mean of 8.2 physiotherapy sessions per year. The Survey includes 115 different resource items which are quantified. The submission then uses accepted sources of unit cost data for cost estimation. The submission cites the unpublished manuscript by Tyas and colleagues[58] to support the use of the cost estimates presented by EDSS state. However we have not been able to reconcile the cost estimates in these two sources, especially for the NHS and PSS perspective. The health state costs (EDSS states) used in the ScHARR model have not been published, so we are unable to compare across analyses. However, as with EDSS utility data (above), estimates of cost by EDSS state are presented by Tappenden and colleagues (2001) as part of the ScHARR model reporting. We present these cost estimates from Tappenden and colleagues in Appendix 7 (p. 127), but we are not aware of the basis for these cost estimates (derived from data from the MS Trust) or the decision making perspective to which they refer .

Table 15: Costs associated with different disease and patient characteristics under different perspectives (UK MS Survey 2005, from submission, Table 8)

		Annual Cost		
Category	**Sub-category**	**NHS & PSS (£)**	**Governmental (£)**	**Societal (£)**
State	EDSS 0	638	2682	16 541
	EDSS 1	927	3242	17 949
	EDSS 2	883	4288	23 176
	EDSS 3	2758	6849	28 958
	EDSS 4	1756	4753	22 657
	EDSS 5	2543	7452	30 598
	EDSS 6	3146	8604	32 166
	EDSS 7	7384	14 217	39 322
	EDSS 8	17 370	27 153	52 686
	EDSS 9	16 307	26 439	52 039
Type	RRMS	†	†	†
	SPMS	56	789	2916
Relapse	No Relapse	†	†	†
	Cost per relapse	228	398	572
Gender	Female	†	†	†
	Male	0	100	1577
DMT (IFN-beta)	No Treatment	†	†	†
	IFN-beta Treatment	8652	8652	8652
DMT (GA)	No Treatment	†	†	†
	GA Treatment	6202	6202	6202
DMT by EDSS State (IFN-beta)	With DMT in EDSS 0-2	†	†	†
	With DMT in EDSS 3-6	236	236	236
DMT by State EDSS (GA)	With DMT in EDSS 0-2	†	†	†
	With DMT in EDSS 3-6	-587	-587	-587
Age	Age	0	-49	-318

* P < 0.01. Reference case (refers to the reference case in the economic evaluation presented in [Manufacturer Submission]).

† = reference case. DMG = Direct Medical cost funded by Government. DNMG = Direct Non-Medical cost funded by Government.

The table reports the profile of the direct costs of managing MS in the UK. These costs were collected in the UK MS Survey 2005 (see Manufacturer Submission]).

Where the manufacturer submission uses a societal decision-making perspective it includes costs associated with NHS and PSS perspective, government costs, out of pocket expenses for patients, and indirect costs, including an estimate of costs associated with loss of income (estimated at approximately £1,500 per month).

Resource use and costs associated with adverse events

The analysis includes an estimate of the cost associated with investigation of patients suspected of PML when on natalizumab treatment, and costs associated with testing for the presence of natalizumab anti-bodies (NAB). It assumes that 23% of people on natalizumab treatment will be tested for NAB, with 20.1% of those tested needing a second test (NAB testing assumed for year 1 only). This assumption is based on AFFIRM data, and may be higher than that expected in clinical practice (where treating clinicians may not use NAB testing for treatment decisions).

The costs associated with PML investigation are based on an expected 9% of people requiring a MRI test, with 1.1% of these requiring a subsequent cerebrospinal fluid (CSF) test. The costings for these patients are estimated at a mean cost of £313 per brain MRI, plus a small additional amount for mean subsequent costs. However, the model in this instance is somewhat perverse as it assumes that all patients undergoing PML testing will miss out on 1-month of natalizumab treatment, a cost saving of over £1,130, and therefore where patients are investigated for PML it is 'cost-saving' with no detrimental consequences (no disutility, no cases of PML diagnosed). The impact of this issue (proportion undergoing PML tests) has a negligible impact on the cost per QALY, but where the proportion tested increases the cost per QALY falls. In practice it is expected that an MRI scan will be available quite quickly (within 10-days), for the majority of patients, and there will be no need to withdraw patients from natalizumab treatment for 1-month.

A further issue related to concerns over PML is that, in practice, all patients treated with natalizumab may undergo a baseline MRI scan, in order to consider any future concern over PML. This cost is not presently included in the analysis for all patients. Where these PML issues are altered (no drug withdrawal, MRI scan for all at start) it would lead to an additional on-cost for natalizumab in the region of £1,300, which would have an impact on the cost effectiveness estimate (we find only a small increase in the cost per QALY, see additional PenTAG sensitivity analysis). Given the increased awareness of PML monitoring when on treatment with natalizumab, it may also lead to additional clinical management costs (more consultant/clinical time).

The model estimates treatment costs for three adverse events which are included in the model – hypersensitivity, urticaria, anaphylactic reaction. The costs for hypersentitivity and urticaria are relatively small (£10 to £70), but anaphylactic reaction is estimated to cost £471 to treat. The cost estimates for these SAEs seem reasonable. The model assumes a mean annual costs associated with natalizumab AEs of £7.24 and £2.46 per patient in first and subsequent years respectively.

5.3.4 Consistency

We have examined the submitted Excel model for internal and external consistency and accuracy.

Internal consistency

The manufacturers submission reports that checking of the internal consistency of the model has been undertaken, reporting testing in Table 81 of the submission, and also reporting independent peer review on a number of aspects of the model.

We have undertaken extensive checking of the Excel programming and mathematical logic of the model, and find the model to be well set out and accurate (with the exception of the items listed below). All equations in the model have been checked for internal consistency/accuracy. However, we have not undertaken a full 'checking' process against all cells in the model. The model is fully executable and we have been able to replicate CEA results presented in the submission, the sensitivity analysis presented (in almost all scenarios), and the probabilistic sensitivity analysis presented. The model does allow user inputs, and inputs changed in the '*INPUT Options*' worksheet (cells C7:C18, E26), '*INPUT Cohort*' sheet (cells F3:F7), and '*INPUT AEs*' sheet (cell C5) produce immediate changes to the deterministic results in the '*INPUT Options*' sheet (cells B28:G36). The '*INPUT Options*' sheet worksheet is clear and user friendly. The model is shown using the baseline scenarios described in the submission, and the user can choose the patient subgroups described in the submission (against treatment strategies for the RES and SOT treatment groups).

We have considered the mathematical derivation of transit probabilities from data presented by the AFFIRM trial (submission Tables 40 & 41), and relapse rates derived from data presented from the AFFIRM RCT (submission Tables 46 & 47). Whilst unable to replicate the exact methods used in the submission, we have been able to estimate transit probabilities and relapse rates which are very similar to those presented in the submission, from the data available, and find the mathematical methods used appropriate. We have been unable to check the accuracy of the data used from the London Ontario dataset as the dataset is not available in the public domain, and is not presented in the manufacturer submission (other than via the transit probabilities used in the spreadsheet model).

When checking the programming of the Excel model, we discovered several errors in the early stages of the review process;

- In worksheet 'DSS', cells AA87:93 made incorrect references. This error affected the disability progression probabilities in the RES subgroup, and led to an underestimate of the cost per QALY estimates for this subgroup.
- The formula for discounting costs and benefits in worksheet 'CALC Discounting' was incorrect (this had a minor impact on cost-effectiveness results).
- The cost of a relapse was incorrectly referenced in sheet 'CALC Costs' (this had a minor impact on cost-effectiveness results).

The above observations were made to NICE at the earliest opportunity, and the manufacturer produced corrected results in an addendum (dated 22/12/06) to their original submission to NICE. This report refers to these corrected cost-effectiveness results.

In later examination of the model, we also report (for completeness) the following observation/error:

- In sheet 'CALC Costs', cells BC10:59, it is assumed that NAB testing costs are incurred at the start of each year.
- But this is inconsistent with the approach stated in the report, where the submission states that NAB testing costs are incurred only at the start of the first year.

However, we note that this potential error has only a minor impact on cost-effectiveness results.

External consistency

The manufacturer submission reports that the external consistency of the model has been considered. In section 6.2.12.3 of the submission the issue of validity is considered in the context of (i) external independent review, (i) predictions of the model compared to data inputs, and (iii) consideration of model outputs against other studies of a similar nature. In the earlier section of the submission (5.8.3) there is also some consideration of the fit of the model compared to data inputs, in the context of calculating and adjusting transition probabilities.

The submission states that the methods used were valid (Section 6.2.12.3). The submission compares AFFIRM data to the multi-state-model (MSM) used to derive transit probabilities, presenting evidence that the predictive power of the transit probabilities derived from the MSM model is high (submission Section 5.8.3). The data presented in the submission in Figures 11 & 12, shows a similar proportion of patients (BSC) across EDSS states (at 2-years) for both the MSM data and AFFIRM data. These data, showing the EDSS profile at 2-years, are also consistent with our additional analysis on disability progression, although our analysis suggests that the model is predicting a more rapid rate of disability progression than that reported in the AFFIRM trial (discussed below).

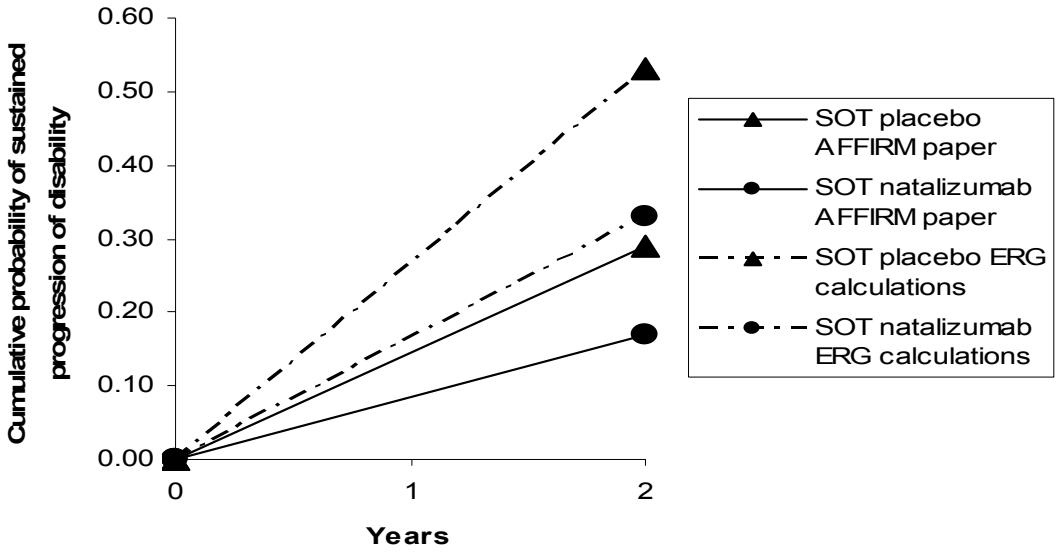
As part of this validity check the submission states that the model was used to predict the endpoint data from the NAT arm of the AFFIRM trial. But no detail is provided here. The submission reports that the results from this validation exercise were used to adjust the level of risk reduction (treatment effect) applied to transition probabilities in the RRMS to SPMS strata of the model (fitting trial data to the model structure).

We have concerns over the different rate of disability progression predicted in the manufacturer model and that presented in the AFFIRM trial against the primary endpoint of AFFIRM. We have undertaken additional analysis to compare the prediction of the model against the endpoint of the AFFIRM trial.

The AFFIRM trial uses as its primary endpoint the "cumulative probability of sustained progression of disability, … defined as an increase of 1.0 or more on the EDSS from a baseline score of 1.0 or more or an increase of 1.5 or more from a baseline score of 0 that was sustained for 12-weeks (progression could not be confirmed during a relapse)" (Polman and colleagues, 2006, p901). Polman and colleagues (2006), in the publication of AFFIRM data, present Kaplan-Meier plots (their Figure 2) of the time to sustained progression of disability among patients receiving natalizumab (ITT group, SOT subgroup) compared to placebo. These are the main clinical findings presented from the AFFIRM trial; a cumulative probability of progression of 17% in the natalizumab group compared to 29% in the placebo group, over 2-years (an absolute risk reduction of 12%; hazard ratio of 0.58 for natalizumab). The model used in the manufacturer model applies this form of data to define the relative risk reductions applied in the model for natalizumab. The submission uses the primary endpoint with sustained progression for 24-weeks (See submission Table 18); applying a hazard ratio of 0.46 for the SOT subgroup and 0.36 for the RES subgroup.

To compare the findings from the AFFIRM trial with the disability progression predicted in the manufacturer model, we applied the transition matrices presented in the model (worksheet: 'Transit Matrices') to the baseline cohort distribution used in the model across EDSS health states, over a period of 2-years (i.e. as per endpoint presented in AFFIRM). We plotted the transitions in the BSC and natalizumab cohorts over 2 x 1-year model cycles (i.e. progression at 2 years) and counted the numbers of the cohort who progressed by 1.0 EDSS point or more (mortality over this 2-year period was also included). We note that 6% of the starting MS cohort are in EDSS state 0 and sustained progression for these people is based on a move of 1.5 EDSS points or more, however there are very small numbers involved here, and for simplicity our analysis considers a move of 1.0 point or more. Our analysis estimates the cumulative probability of sustained progression of disability in the model at 2-years. Figure 3 reports findings against the SOT group from model predictions for BSC and natalizumab treated cohorts, and plots model predictions against the results presented for the AFFIRM trial (ITT/SOT group). The AFFIRM data plots reflect the data presented by Polman and colleagues (2006),[59] 17% in natalizumab group versus 29% in the placebo group (extracted from Polman and colleagues, Figure 2[60]).

Figure 3: Comparison of disability progression predicted in the model versus data reported from AFFIRM; cumulative probability of sustained progression of disability for the SOT/ITT patient subgroup at two years.



Our analysis of the manufacturers model shows a much higher cumulative probability of sustained disability progression (at 2-years) for the SOT subgroup (ITT group of AFFIRM) than in the AFFIRM trial (in Table 18 of the manufacturer's submission). Results show a cumulative probability of progression of 33% in the natalizumab treated group versus 53% for BSC (absolute risk reduction of 20%). A similar conclusion applies for the RES subgroup.

This analysis is simple, and is based on 1-year cycles (transit probabilities), but it clearly indicates that the rate of disability progression predicted in the model is much greater than that reported in the AFFIRM trial. It also indicates that the treatment effect predicted for natalizumab in the model through using relative risk reductions in each cycle is much greater than that reported for this endpoint by the AFFIRM trial.

Given that natalizumab is predicted to both reduce the risk of disability progression, and the costs and outcome consequences (utility decrements) associated with disability progression, we suggest that predicting a greater rate of disability progression presents a more favourable treatment effect (greater level of disability prevented) and a potential for underestimating the cost per QALY associated with treatment (e.g. larger utility losses can be avoided by delaying progression of disability).

We note that the submission reports an estimated annual rate of disability progression from the MSM transition probabilities for the placebo arm (BSC) ITT population of the AFFIRM trial (p97). The mean annual rate of disability progression is estimated to be 0.27 EDSS states (95% CIs: 0.11, 0.43) for the ITT (SOT) subgroup; for the RES subgroup the mean annual rate is 0.46 (0.16, 0.79).

5.3.5 Assessment of uncertainty

One-way sensitivity analyses

The manufacturer submission and subsequent addendum (dated 22/12/06) present extensive sensitivity analyses on the key parameters in the model (Table 85 of report). We have checked all of these univariate-sensitivity analyses, and agree with all except two. In the two cases shown in Table 16 we find different cost per QALY data. In one instance (relapse rates) the differences are very small, and unimportant in the context of the broader report. For sensitivity analysis against the use of the London Ontario data for progression parameters we find quite different results, especially so in the comparison of natalizumab and BSC.

Table 16: Checks against manufacturer univariate sensitivity analysis.

	RES BSC	RES IFN-β	RES GA	SOT BSC	SOT IFN-β	SOT GA
Base case	**44.6**	**32.0**	**34.6**	**56.1**	**43.4**	**44.3**
Relapse duration -50%:						
- manufacturer submission	0.2	0.2	0.1	0.1	0.1	0.1
- PenTAG calculations	0.4	0.3	0.3	0.3	0.3	0.2
Progression data from London Ontario data:						
- manufacturer submission	23.2	10.3	15.9	27.8	11.9	19.3
- PenTAG calculations	7.1	13.5	20.5	3.6	16.6	26.0

Figures in bold are ICERs in £,000s. Figures not in bold are changes in ICER from base case, in £,000s

Whilst the manufacturer submission reports extensive one-way sensitivity analysis, for completeness we have performed additional sensitivity analyses, against mortality rates and drug costs, and around assumptions in the model over withdrawal, disutility associated with treatment, and costs for PML investigations, see Table 17.

Table 17: PenTAG Univariate sensitivity analyses. Figures in bold are ICERs in £,000s. Figures not in bold are changes in ICER from base case, in £,000s.

	RES BSC	RES IFN-β -beta	RES GA	SOT BSC	SOT IFN-β -beta	SOT GA
Base case	**44.6**	**32.0**	**34.6**	**56.1**	**43.4**	**44.3**
Mortality rates zero	-1.6	-0.9	-1.3	-1.8	-1.0	-1.4
Mortality rates, base case -50%	-0.8	-0.5	-0.6	-0.9	-0.5	-0.7
Mortality rates, base case x 2 (doubled)	1.5	0.9	1.2	1.7	1.0	1.4
Cost of NAT -10%	-5.5	-7.7	-6.4	-6.6	-10.1	-7.8
Cost of NAT -15%	-8.3	-11.6	-9.5	-9.8	-15.1	-11.7
Cost of NAT -25%	-13.7	-19.2	-15.8	-16.3	-25.1	-19.4
Cost of NAT +25%	13.7	19.2	15.8	16.3	25.1	19.4
Cost of IFN-β -25%	-	9.7	-	-	13.6	-
Cost of GA -25%	-	-	5.2	-	-	6.8
For 1st year only, disutility for 30% of patients on treatment (NAT, IFN-β and GA) is 0.05	0.2	1.0	0.1	0.3	1.8	0.2
Withdrawal rate of 10% in first year, 3% all later years for all drugs	1.6	2.4	2.3	2.0	2.8	2.8
All natalizumab patients initially receive an MRI, no drug withdrawal (for PML investigation)	0.4	0.6	0.5	0.5	0.8	0.6

On the basis of the above data, showing that the ICER is sensitive to the cost of natalizumab, we have considered threshold analysis on the cost of natalizumab required that would dictate a cost per QALY of £30,000, holding all other base case assumptions constant. For the RES cohort, comparing natalizumab with BSC, IFN-β or GA the monthly cost of natalizumab would need to be £829 (-27%), £1,100 (-3%) and £1,049 (-7%) respectively, to present at a level of £30,000 per QALY. For the SOT cohort, comparing natalizumab with BSC, IFN-β or GA the monthly cost of natalizumab would need to be £676 (-40%), £979 (-13%) and £921 (-18%) respectively, to present at a level of £30,000 per QALY.

Probabilistic sensitivity analysis

The manufacturer's model includes a probabilistic sensitivity analysis (PSA), presented as cost-effectiveness acceptability curves (CEACs) and a scatter-plot in the cost-effectiveness plane (submission report p162-167, and in Addendum).

Summary of assumptions for manufacturer's PSA

The methods used for PSA appear to be appropriate, with the exception of the issues noted below which have little effect on the PSA results.

- The distribution of the initial cohort by EDSS follows a Dirichlet distribution.
- The disutility per patient due to administration of treatment (natalizumab, IFN-β or GA) follows a beta distribution. We suggest that the proportion of patients who experience a disutility could also be modelled as a beta distribution.
- The relative relapse rate and progression rate of natalizumab versus placebo and comparators versus placebo follow lognormal distributions.
- The transition probabilities within SPMS states follow a Dirichlet distribution. The model assumes 925 observations for each EDSS state, which comes from the London Ontario data, assuming an equal number of observations per state. Transitions probabilities between EDSS states for RRMS for both the RES and SOT subgroups also follow a Dirichlet distribution.
- The transition probabilities for RRMS to SPMS follow a beta distribution. We found errors in the calculation of the parameters for this distribution. However, these have little effect on PSA results.
- The cost coefficients from the seemingly unrelated regression analysis follow a multivariate normal distribution. The Excel model has hard coded the corresponding mean and covariance matrix, therefore we are unable to check the origin of the data.
- The EDSS health state utilities, with coefficients for type of MS (RRMS or SPMS), time since diagnosis and whether there has been a recent relapse, follow a multivariate normal distribution. As above, the corresponding mean and covariance matrix are hard coded in the Excel model, therefore we are unable to check the origin of the data.

5.4 Summary of uncertainties and key issues

In general the approach taken to model MS and the cost-effectiveness of natalizumab compared to BSC, IFN-β and GA, is a pragmatic one, using available data and drawing heavily on the previous approach to modelling MS in the context of the NICE appraisal process.[48]

The above review of the economic analysis presented in the manufacturer submission provides an outline of the methods used and a critical review of the CEA, modelling methods, and the checking of the Excel spreadsheet model used to assess cost-effectiveness.

We raise a number of concerns, some general and some specific:

- The model is structured around the use of the EDSS disability scale. Whilst this is consistent with previous modelling of MS, undertaken for NICE, it is important to consider the limitation of the EDSS to capture disability, and the broader HRQL associated with MS.

- The model structure allows backward (improving) transition probabilities when people are in RRMS health states EDSS 0-6, and the rationale for this is not clear.

- The model structure uses a reduced number of EDSS health states, banding states together, and the impact of this is not discussed/explored in the model.

- The RES and SOT subgroups are used for the CEA, and the definition of these subgroups in practice involves some uncertainty. SOT subgroup analysis uses the ITT results from AFFIRM (RRMS).

- For IFN-β and GA the same relative risks are applied to modify the disability progression in RES and SOT analyses (differences between IFN-β and GA, but no differences by subgroup).

- The appropriateness of using data from the London Ontario dataset in a HARRMS treatment group should be considered.

- The model of MS for the control group applies data from a number of sources, and data from AFFIRM is from small samples (e.g. RES n=61), over a 2-year time-frame.

- The data from the ITT group of the AFFIRM RCT are used to estimate transition probabilities for the SOT subgroup.

- Data from multiple sources is used to estimate relapse rates by EDSS state (self-report survey data from UK MS Survey 2005, plus observational data from a small epidemiological study, and data from AFFIRM).

- The model applies a constant treatment effect over the 20-year time horizon and there is an absence of evidence to support this assumption.

- We suggest that the model predicts a more rapid rate of disability progression than that reported in the AFFIRM RCT, and this will impact on the cost-effectiveness analysis.

- The UK MS Survey 2005 is used to populate the model (e.g. costs, QALYS, relapse data) and we have concerns over the generalisability of the data from the survey to the MS treatment group in the current (NICE) decision problem.

- The data used to set the disutility parameters (health state disutilities, from Prosser 2003) for treatment options is from small convenience samples, and we believe it is not robust enough to identify the differences used by treatment in the model.

- The assumptions in the model around testing for PML introduce a cost-saving scenario when people are investigated for PML, through temporary non-use of natalizumab, and this may not be the case in practice.

- Use of caregiver disutility estimates in base case analysis (NHS & PSS perspective).

- The submission states that testing for NAB is in year one only, as the development of NAB is likely to occur early on in the treatment period. However, in practice NAB testing may be required in subsequent years.

6 CONCLUSIONS

6.1 Summary of main issues for clinical effectiveness

- Based on one RCT natalizumab is effective at reducing sustained disability progression and relapse rates for people with RRMS compared to placebo.
- Similar results were seen in a subgroup of people with RES in this trial.
- Evidence for people with SOT MS is lacking and has been assumed to be represented by trials undertaken with different populations or combination treatment regimens.
- No head to head trials of natalizumab compared to active DMT have been undertaken and so assessment of effectiveness is based on indirect comparison.
- [redacted]
- [redacted]
- Relapse rates are statistically significantly reduced with natalizumab in people with RES compared to both IFN-β and GA.

6.2 Summary of main issues for cost-effectiveness

- In general the approach taken to model MS and the cost-effectiveness of natalizumab compared to BSC, IFN-β and GA, is a pragmatic one, using available data and drawing heavily on the previous approach to modelling MS in the context of the NICE appraisal process.
- The Kurzke EDSS disability scale is crucial to the model structure. Although widely used, this scale has limitations.
- The model structure allows people to experience improved disability ratings and the rationale for this is not clear. Relapse is accounted for separately.
- The model structure uses a reduced number of EDSS health states and the impact of this is not discussed/explored in the model.

- The appropriateness of using data from the London Ontario dataset in a HARRMS treatment group is not clear.
- The model of MS for the control group applies data from a number of sources, and data from AFFIRM is from small samples (e.g. RES n=61), over a 2-year time-frame.
- The data from the ITT group of the AFFIRM RCT are used to estimate transition probabilities for the SOT subgroup.
- Data from multiple sources (a self-report survey data from UK MS Survey 2005, plus observational data from a small epidemiological study, and data from AFFIRM) is used to estimate relapse rates by EDSS state.
- The model applies a constant treatment effect over the 20-year time horizon and there is an absence of evidence to support this assumption.
- Additional analysis we undertook suggests that the model predicts a more rapid rate of disability progression than that reported in the AFFIRM RCT, and this will impact on the cost-effectiveness analysis.
- The UK MS Survey 2005 is used to populate a number of model parameters (e.g. costs, QALYS, relapse data) and we have concerns over the generalisability of the data from the survey to the MS treatment group in the current decision problem.
- The data used to set the disutility parameters (health state disutilities, from Prosser 2003) for treatment options is from small convenience samples, we believe it is not robust enough to identify the differences used by treatment in the model.
- The model assumes some costs are saved due to investigations because natalizumab is temporarily suspended. This may not be the case in practice.
- Use of caregiver disutility estimates in base case analysis (NHS & PSS perspective).

6.3 Suggested research priorities

- The effectiveness of natalizumab in the SOT group should be investigated through clinical trials.
- Trials of natalizumab compared with IFN-β and GA, and other DMT for MS should be undertaken.
- Long term follow up of DMT effectiveness and safety is needed.
- A better understanding of disability progression patterns in RRMS is needed in order to model the impact of treatments in the long term.
- A better understanding of the value that people with MS place on reductions in relapse rate is needed.

7 APPENDICES

7.1 Appendix 1: Quality assessment of evidence synthesis

Data extraction and quality assessment of STA submission: Natalizumab review

Reference	
Author: Biogen Idec (manufacturer of natalizumab) and Heron Evidence Development (consultancy)	
Research Question	
Aim (Question): Not stated directly in the submission. The scope aims "To appraise the clinical and cost effectiveness of natalizumab in its licensed indications for the treatment of multiple sclerosis" Licensed indications are Rapidly Evolving Severe (RES) Relapsing Remitting Multiple Sclerosis (RRMS) and sub-optimally treated people with RRMS.	
Search strategy (to be undertaken by the information scientist)	
Databases searched: None Search terms: Not stated Limits: None stated Dates: Not stated Critique of search strategy: No formal search was undertaken for natalizumab in the submission. This is justified because the manufacturer believe that they have managed all studies conducted to date.	
Inclusion Criteria of the submission (add comment where possible)	

Because there is no formal search strategy, no formal inclusion criteria are listed, however, relevant studies are listed in table 11 and the following may be assumed:

Study design: RCT

Interventions: Natalizumab IV infusion (alone or in combination with IFN-β)

Population: Adults with RRMS

Primary Outcome measures: rate of clinical relapses at one year, sustained progression of disability at 2 years, brain lesion activity as measured by MRI, adverse effects.

Were any limits placed on inclusion relating to the quality of the RCTs? N/A

Was setting used as an inclusion criteria? N/A

Does the inclusion reflect the information in the decision problem, including the licensed indication? the licensed indication is for RES and SOT populations while trial data is on RRMS.

Are there any known additions that could have been made to assess the technology? No.

Note that information in the report is based on both published data and that held on file by the company, notably the RES subgroup. Some of this data is commercial in confidence and cannot be checked by us.

Results	

Quantity of included studies:

How many RCTs were included: Four.

Were any studies included that do not meet the inclusion criteria: N/A as no inclusion criteria. However, two RCTs have only 12 week follow up and so do not contribute to the assessment of efficacy, providing only some additional safety data. One other RCT assesses the efficacy of natalizumab in addition to IFN-β compared to IFN-β alone which is not the licensed use. The remaining RCT contains a RRMS population and post-hoc analyses is provided about a RES subgroup.

Were there any studies not included that meet the inclusion criteria: Probably not. One RCT of natalizumab combined with GA, compared to GA alone in those with RRMS still experiencing relapses despite treatment with GA, was identified but nit included.

Quality assessment of included studies.

Were the NICE criteria used or an alternative: NICE criteria and Jadad score applied

Have all questions been applied by the manufacturer. Yes

Are there any discrepancies with our assessment of quality and the manufacturers: yes – our assessment of generalisability of the trial data to clinical UK population is more cautious than that by the manufacturer.

How was the quality assessment applied by the manufacturer: Descriptive only.

Method of analysis: qualitative

Is data tabulated: Yes

Does the narrative reflect the data in the tables: Yes

Method of analysis: quantitative

What method of meta-analysis was stated as being applied by the manufacturer: Only safety data was meta-analysed to facilitate indirect comparison with IFN-β and GA. Full details are not given.

What factors were taken into account in a sensitivity analysis: None undertaken

Do these appear to be reasonable: NA

What was the combined treatment effect: (Should include point estimates and confidence intervals/standard deviations, P values etc). Treatment effect not pooled. Main results for ITT population and post-hoc RES subgroup shown below.

ITT Population **Outcome**	**Natalizumab (n = 627)**	**Placebo (n = 315)**	**Absolute risk reduction**	**Hazard ratio (95% CI)**
Probability of sustained disability progression (defined as an increase in EDSS sustained for 12 weeks) at two years †	0.17	0.29	0.12	0.58 (0.43, 0.77)
Probability of sustained disability progression (defined as an increase in EDSS sustained for 24 weeks) at two years † ‡ §	0.11	0.23	0.12	0.46 (0.33, 0.64)
Annualised relapse rate at one year	0.26	0.81	0.55	0.68 (0.59, 0.74)
Annualised relapse rate at two years §	0.24	0.73	0.50	0.68 (0.60, 0.74)
RES Subgroup **Outcome**	**Natalizumab (n=148)**	**Placebo (n=61)**	**Absolute risk reduction**	**Hazard ratio (95% CI)**
Probability of sustained disability progression (defined as an increase in EDSS sustained for 12 weeks) at two years †	0.14	0.29	0.15	0.47 (0.24, 0.93)
Probability of sustained disability progression (defined as an increase in EDSS sustained for 24 weeks) at two years † ‡ §	0.10	0.26	0.16	0.36 (0.17, 0.76)
Annualised relapse rate at two years §	0.28	1.46	1.17	0.81 (0.70, 0.88)

Selected AEs were pooled as shown below (those that were reported by both the nAT trial and the GA or IFN-B systematic reviews) listed below

Endpoints from NAT trials (n)	AFFIRM, MS 201, MS231: NAT vs. placebo			
	RR	lcl	ucl	p
Abdominal discomfort (1152)	1.96	0.22	17.69	0.55
AEs causing treatment withdrawal (1224)	1.54	0.85	2.81	0.16
Anxiety (1152)	0.70	0.44	1.11	0.13
Appetite decreased NOS (939)	0.50	0.13	1.98	0.32
Constipation (1224)	0.88	0.55	1.40	0.58
Muscle contraction, involuntary (1152)	0.80	0.10	6.44	0.83
Dizziness (1224)	0.94	0.67	1.32	0.71
Somnolence (1152)	2.46	0.63	9.55	0.19
Dyspnoea (1152)	1.13	0.47	2.73	0.79
Syncope (1152)	0.77	0.22	2.67	0.68
Headache (1224)	1.07	0.91	1.26	0.44
Infusion site erythema (939)	1.49	0.16	14.29	0.73
Joint stiffness (939)	0.83	0.20	3.45	0.80
Nausea (1224)	0.65	0.32	1.33	0.24
Infusion site pain (1152)	2.16	0.33	14.20	0.42
Convulsions (1152)	0.67	0.13	3.57	0.64
Rash NOS (1152)	1.21	0.80	1.83	0.37
Infusion site pruritus (939)	1.00	0.09	10.93	1.00
Infusion site swelling (1152)	1.23	0.16	9.27	0.84
Vomiting (939)	0.65	0.42	1.02	0.06
Suicidal ideation (939)	2.49	0.12	51.75	0.56
Fatigue (1011)	1.62	0.80	3.28	0.18
Pyrexia (1152)	1.49	0.34	6.63	0.60
Influenza-like illness (1011)	0.79	0.50	1.25	0.32
Infusion site reaction (1152)	2.29	0.59	8.96	0.23
Psychiatric disorders (939)	0.57	0.21	1.55	0.27
Myalgias / Arthralgia (939)	1.26	0.95	1.68	0.11

Was there an assessment of heterogeneity: Not stated

Adverse effects.

Are the most common adverse events reported or is there any selection of reporting: Serious adverse effects are listed only if occurring in at least two patients receiving natalizumab. Most data comes form the published trial and it is possible there is more held on file. Note that the indirect comparison of natalizumab and GA or IFN-B trials only reports those recorded in the Cochrane reviews of these comparators an therefore excludes the only 2 AEs that are significantly more common in natalizumab than placebo – allergic reaction (9% vs 4%, p=0.012) and fatigue (27% vs 21%, p=0.048).

Interpretation of evidence	
Is the interpretation of the manufacture justified: Yes although there is some confusion between assessments undertaken in the clinical review section of the report, where indirect comparisons are undertaken and data form the SENTINEL trails is taken as a proxy for effectiveness on the SOT population, and subsequent modelling, where indirect comparison data is not used and the SOT population is modelled using the placebo arm of the AFFIRM trial.	
General comments	
Are there any differences in baseline characteristics of patients and controls? Not in the ITT population although in the RES subgroup the natalizumab arm is slightly younger than the placebo arm (mean 33.7 vs 36.4 p=0.038) although this is probably not clinically significant. Disease duration is also slightly shorter in the natalizumab arms at a median of 4.0 years compared to 5.0 years in the placebo arm (p=0.501) Note that the trial is not powered to detect differences in the subgroup (n=209, compared to 942 in the ITT population.)	

7.2 Appendix 2: Search strategies used to identify trials of natalizumab and of comparators

Clinical effectiveness searches for natalizumab in RRMS Searched 21-12-06

Databases and years searched	Date searched and search files	Number retrieved	Number of hits (download file)
Cochrane Library – CENTRAL – Issue 4/2006	#1 MeSH descriptor Multiple Sclerosis explode all trees #2 (natalizumab or tysabri or antegren) #3 (#1 AND #2)	5	5
Ovid MEDLINE(R) <1966 to November Week 3 2006>	1 natalizumab.mp. (182) 2 tysabri.mp. (18) 3 antegren.mp. (14) 4 1 or 2 or 3 (185) 5 exp Multiple Sclerosis/ (28586) 6 Multiple Sclerosis, Relapsing-Remitting/ (1262) 7 (multiple sclerosis or ms).ti,ab. (109820) 8 5 or 6 or 7 (114354) 9 4 and 8 (116) 10 limit 9 to english language (104) 11 randomized controlled trial.pt. (242391) 12 controlled clinical trial.pt. (78115) 13 exp Randomized Controlled Trials/ (50346) 14 exp Random Allocation/ (59887) 15 double-blind method/ (93234) 16 Single-Blind Method/ (11141) 17 11 or 12 or 13 or 14 or 15 or 16 (411230) 18 clinical trial.pt. (467833) 19 exp Clinical Trials/ (200143) 20 (clin$ adj9 trial$).ab,ti. (121598) 21 ((singl$ or doubl$ or trebl$) adj9 (blind$ or mask$)).ti,ab. (90466) 22 placebo$.sh. (27716) 23 placebo$.ti,ab. (103741) 24 random$.ti,ab. (380181) 25 research design.sh. (47231) 26 or/18-25 (870985) 27 26 not 17 (488655) 28 comparative study.sh. (1381246) 29 exp Evaluation Studies/ (611098) 30 Follow-Up Studies/ (347849) 31 Prospective Studies/ (225752) 32 (control$ or prospectiv$ or volunteer$).ti,ab. (1799090) 33 or/28-32 (3555936) 34 33 not (17 or 27) (2946566)	67	67

	35 17 or 27 or 34 (3846451) 36 10 and 35 (67)		
EMBASE <1980 to 2006 Week 48>	1 natalizumab.mp. (700) 2 Natalizumab/ (684) 3 tysabri.mp. (172) 4 antegren.mp. (151) 5 1 or 2 or 3 or 4 (714) 6 exp Multiple Sclerosis/ (26489) 7 Multiple Sclerosis, Relapsing-Remitting/ (26489) 8 (multiple sclerosis or ms).ti,ab. (99725) 9 6 or 7 or 8 (105095) 10 5 and 9 (417) 11 limit 10 to english language (352) 12 randomization/ (21013) 13 controlled study/ (2304140) 14 single blind procedure/ (6231) 15 placebo/ (92599) 16 double blind procedure/ (62108) 17 clinical trial/ (404757) 18 crossover procedure/ (18076) 19 placebo$.tw. (95457) 20 blind$ fashion.tw. (3515) 21 random$.tw. (322147) 22 clinical trial?.tw. (95348) 23 or/12-22 (2695278) 24 limit 23 to human (1714858) 25 11 and 24 (271)		271
Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <December 01, 2006>	1 natalizumab.mp. (20) 2 tysabri.mp. (3) 3 antegren.mp. (2) 4 1 or 2 or 3 (21) 5 [exp Multiple Sclerosis/] (0) 6 Multiple Sclerosis, Relapsing-Remitting/ (0) 7 (multiple sclerosis or ms).ti,ab. (6032) 8 5 or 6 or 7 (6032) 9 4 and 8 (11) 10 limit 9 to english language (11) 11 randomized controlled trial.pt. (292) 12 controlled clinical trial.pt. (20) 13 [exp Randomized Controlled Trials/] (0) 14 [exp Random Allocation/] (0) 15 double-blind method/ (0) 16 Single-Blind Method/ (0) 17 11 or 12 or 13 or 14 or 15 or 16 (312) 18 clinical trial.pt. (309) 19 [exp Clinical Trials/] (0)	5	3

	20 (clin$ adj9 trial$).ab,ti. (4653) 21 ((singl$ or doubl$ or trebl$) adj9 (blind$ or mask$)).ti,ab. (1820) 22 placebo$.sh. (0) 23 placebo$.ti,ab. (2528) 24 random$.ti,ab. (18193) 25 research design.sh. (0) 26 or/18-25 (22345) 27 26 not 17 (22033) 28 comparative study.sh. (1) 29 [exp Evaluation Studies/] (0) 30 Follow-Up Studies/ (0) 31 Prospective Studies/ (1) 32 (control$ or prospectiv$ or volunteer$).ti,ab. (61276) 33 or/28-32 (61278) 34 33 not (17 or 27) (51694) 35 17 or 27 or 34 (74039) 36 10 and 35 (5) 37 from 36 keep 1-2,4 (3)		
ISI Science Citation Index 1970-2006 Searched 4-12-06	☐ #10 **52** #9 AND #5 *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #9 **>100,000** #8 OR #7 OR #6 *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #8 **>100,000** TS=random* *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #7 **19,787** TS=(clin* SAME stud*) *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #6 **>100,000** TS=(clin* SAME trial*) *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #5 **136** #4 AND #3 *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006* ☐ ☐ #4 **35,290** TS=(multiple sclerosis) *DocType=All document types; Language=English;* ☐		52

			Database=SCI-EXPANDED; Timespan=1970-2006			
	☐ #3	**281**	#2 OR #1 *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006*	☐		
	☐ #2	**49**	TS=(tysabri or antegren) *DocType=All document types; Language=English; Database=SCI-EXPANDED; Timespan=1970-2006*	☐		
	☐ #1	**257**	TS=natalizumab *DocType=All document types*			
Web of Science Proceedings 1990-present Searched 4-12-06	☐ #3	**36**	#2 AND #1 *DocType=All document types; Language=English; Database=STP; Timespan=1990-2006*	☐		36
	☐ #2	**4,705**	TS=(multiple sclerosis) *DocType=All document types; Language=English; Database=STP; Timespan=1990-2006*	☐		
	☐ #1	**83**	TS=(natalizumab or tysabri or antegren) *DocType=All document types; Language=English; Database=STP; Timespan=1990-2006*			
BIOSIS Previews *1990-2006 Searched 4-12-06*	☐ #3	**34**	#2 AND #1 *DocType=All document types; LitType=Meeting Abstract; Language=English; Taxa Notes=All Taxa Notes; Database=BIOSIS Previews; Timespan=1990-2006*	☐		34
	☐ #2	**4,756**	TS=(multiple sclerosis) *DocType=All document types; LitType=Meeting Abstract; Language=English; Taxa Notes=All Taxa Notes; Database=BIOSIS Previews; Timespan=1990-2006*	☐		
	☐ #1	**65**	TS=(natalizumab or tysabri or antegren) *DocType=All document types; LitType=Meeting Abstract; Language=English; Taxa Notes=All Taxa Notes; Database=BIOSIS Previews; Timespan=1990-2006*			
DARE	As Cochrane					0

NHS EED	As Cochrane	0	0
HTA database	As Cochrane	1	1
Current Controlled Trials including MRC Trials dB http://controlled-trials.com/	(natalizumab or tysabri) and multiple sclerosis		3
Clinical Trials.gov http://clinicaltrials.gov/	(natalizumab or tysabri) and multiple sclerosis		6
	TOTAL refs in database after deduplication		292

Clinical effectiveness searches: Interferon Beta and RRMS Searched 21-12-06

Databases and years searched	**Date searched and search files**	**Number of hits**
Cochrane Library – CDSR – Issue 4/2006	#1 MeSH descriptor Multiple Sclerosis, Relapsing-Remitting explode all trees #2 ((Multiple Sclerosis NEAR\5 Relapsing-Remitting) or RRMS) #3 interferon beta or interferon-beta or Avonex or Rebif or Beta?eron	2
Cochrane Library – CENTRAL – Issue 4/2006	#4 (#1 OR #2) #5 (#3 AND #4), from 2001 to 2006 #6 (glatiramer acetate or Copaxone) #7 (#4 AND #6), from 2004 to 2006	60
Cochrane Library – NHSEED – Issue 4/2006		4
Ovid MEDLINE(R) <1996 to November Week 3 2006> Saved as med-nat-comp-rrms-extras	1 Multiple Sclerosis, Relapsing-Remitting/ (1166) 2 (((Multiple Sclerosis or MS) adj5 Relapsing Remitting) or RR?MS).mp. (2153) 3 1 or 2 (2153) 4 Clinical Trial.pt. (227082) 5 Randomized Controlled trial.pt. (127963) 6 Multicenter Study.pt. (60362) 7 Controlled Clinical Study.pt. (0) 8 clinical studies.me. (0) 9 Cross-Over Studies.me. (16573) 10 Single-Blind Method.me. (7816) 11 Double-Blind Method.me. (43880) 12 Random Allocation.me. (21524) 13 Follow-Up Studies.me. (161397) 14 Prospective Studies.me. (135417) 15 Placebos.me. (7716)	207

	16 (placebo$ or multicentr$ or comparative study or comparative studies).mp. (631534) 17 (random$ or clinical study$).mp. (299165) 18 (single or double or treble or triple).mp. (435276) 19 (mask$ or blind$ or cross over or crossover or follow up).mp. (385557) 20 18 and 19 (87950) 21 4 or 5 or 6 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 20 (1093273) 22 3 and 21 (953) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (3477) 24 (glatiramer acetate or Copaxone).mp. (358) 25 22 and 23 (367) 26 limit 25 to english language (324) 27 limit 26 to yr="2001 - 2007" (247) 28 22 and 24 (90) 29 limit 28 to (english language and yr="2004 - 2007") (33) 30 (comment or letter or editorial or review).pt. (1101345) 31 27 not 30 (207) 32 29 not 30 (25)	
Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <December 20, 2006>	1 Multiple Sclerosis, Relapsing-Remitting/ (0) 2 ((Multiple Sclerosis adj5 Relapsing-Remitting) or RRMS).mp. (45) 3 1 or 2 (45) 4 Clinical Trial.pt. (311) 5 Randomized Controlled trial.pt. (306) 6 Multicenter Study.pt. (16) 7 Controlled Clinical Study.pt. (0) 8 [clinical studies.me.] (0) 9 [Cross-Over Studies.me.] (0) 10 [Single-Blind Method.me.] (0) 11 [Double-Blind Method.me.] (0) 12 [Random Allocation.me.] (0) 13 [Follow-Up Studies.me.] (0) 14 [Prospective Studies.me.] (0) 15 [Placebos.me.] (0) 16 (placebo$ or multicentr$ or comparative study or comparative studies).mp. (4704) 17 (random$ or clinical study$).mp. (20445) 18 (single or double or treble or triple).mp. (42388) 19 (mask$ or blind$ or cross over or crossover or follow up).mp. (18036) 20 18 and 19 (3663) 21 4 or 5 or 6 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 20 (24589) 22 3 and 21 (9) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (84) 24 (glatiramer acetate or Copaxone).mp. (26) 25 22 and 23 (3) 26 limit 25 to english language (3) 27 limit 26 to yr="2001 - 2007" (3) 28 22 and 24 (2) 29 limit 28 to (english language and yr="2004 - 2007") (2)	3

	30 (comment or letter or editorial or review).pt. (26645) 31 27 not 30 (3)	
EMBASE <1996 to 2006 Week 50>	1 random$.ti,ab. (221123) 2 factorial$.ti,ab. (4610) 3 (crossover$ or crossover$ or cross-over$).ti,ab. (18573) 4 placebo$.ti,ab. (55546) 5 (doubl$ adj blind$).ti,ab. (39320) 6 (singl$ adj blind$).ti,ab. (3538) 7 Crossover Procedure/ (14161) 8 Double Blind Procedure/ (43369) 9 Randomized Controlled Trial/ (98661) 10 Single Blind Procedure/ (5330) 11 or/1-10 (275641) 12 exp ANIMAL/ or NON HUMAN/ or exp ANIMAL EXPERIMENT/ (471846) 13 exp HUMAN/ (3120290) 14 12 and 13 (47560) 15 12 not 14 (424286) 16 11 not 15 (256075) 17 Multiple Sclerosis/ (17742) 18 (((Multiple Sclerosis or MS) adj5 Relapsing Remitting) or RR?MS). 1696 19 17 or 18 (17891) 20 *beta interferon/ or *beta1 interferon/ (1950) 21 beta interferon/ or beta1 interferon/ (5215) 22 (letter or editorial or "review").pt. (916559) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (3220) 24 21 or 23 (7081) 25 16 and 19 and 24 (358) 26 limit 25 to (english language and yr="2001 - 2007") (232) 27 26 not 22 (166) 28 (glatiramer acetate or Copaxone).mp. (755) 29 Glatiramer/ (1263) 30 28 or 29 (1340) 31 16 and 19 and 30 (152) 32 limit 31 to (english language and yr="2004 - 2007") (51) 33 32 not 22 (29) 34 from 27 keep 1-166 (166)	166
	TOTAL refs in database after deduplication Keyworded Include in User def 1 field (separate export)	305 65 Includes

Clinical effectiveness searches: Glatiramer Acetate and RRMS Searched 21-12-06

Databases and years searched	**Date searched and search files**	**Number of hits**
Cochrane Library – CDSR – Issue 4/2006	#1 MeSH descriptor Multiple Sclerosis, Relapsing-Remitting explode all trees #2 ((Multiple Sclerosis NEAR\5 Relapsing-Remitting) or RRMS) #3 interferon beta or interferon-beta or Avonex or Rebif or Beta?eron	0

Cochrane Library – CENTRAL – Issue 4/2006	#4 (#1 OR #2) #5 (#3 AND #4), from 2001 to 2006 #6 (glatiramer acetate or Copaxone) #7 (#4 AND #6), from 2004 to 2006	2
Cochrane Library – NHSEED – Issue 4/2006		1
Ovid MEDLINE(R) <1996 to November Week 3 2006>	1 Multiple Sclerosis, Relapsing-Remitting/ (1166) 2 (((Multiple Sclerosis or MS) adj5 Relapsing Remitting) or RR?MS).mp. (2153) 3 1 or 2 (2153) 4 Clinical Trial.pt. (227082) 5 Randomized Controlled trial.pt. (127963) 6 Multicenter Study.pt. (60362) 7 Controlled Clinical Study.pt. (0) 8 clinical studies.me. (0) 9 Cross-Over Studies.me. (16573) 10 Single-Blind Method.me. (7816) 11 Double-Blind Method.me. (43880) 12 Random Allocation.me. (21524) 13 Follow-Up Studies.me. (161397) 14 Prospective Studies.me. (135417) 15 Placebos.me. (7716) 16 (placebo$ or multicentr$ or comparative study or comparative studies).mp. (631534) 17 (random$ or clinical study$).mp. (299165) 18 (single or double or treble or triple).mp. (435276) 19 (mask$ or blind$ or cross over or crossover or follow up).mp. (385557) 20 18 and 19 (87950) 21 4 or 5 or 6 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 20 (1093273) 22 3 and 21 (953) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (3477) 24 (glatiramer acetate or Copaxone).mp. (358) 25 22 and 23 (367) 26 limit 25 to english language (324) 27 limit 26 to yr="2001 - 2007" (247) 28 22 and 24 (90) 29 limit 28 to (english language and yr="2004 - 2007") (33) 30 (comment or letter or editorial or review).pt. (1101345) 31 27 not 30 (207) 32 29 not 30 (25)	25
Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <December 20, 2006>	1 Multiple Sclerosis, Relapsing-Remitting/ (0) 2 ((Multiple Sclerosis adj5 Relapsing-Remitting) or RRMS).mp. (45) 3 1 or 2 (45) 4 Clinical Trial.pt. (311) 5 Randomized Controlled trial.pt. (306) 6 Multicenter Study.pt. (16)	2

	7 Controlled Clinical Study.pt. (0) 8 [clinical studies.me.] (0) 9 [Cross-Over Studies.me.] (0) 10 [Single-Blind Method.me.] (0) 11 [Double-Blind Method.me.] (0) 12 [Random Allocation.me.] (0) 13 [Follow-Up Studies.me.] (0) 14 [Prospective Studies.me.] (0) 15 [Placebos.me.] (0) 16 (placebo$ or multicentr$ or comparative study or comparative studies).mp. (4704) 17 (random$ or clinical study$).mp. (20445) 18 (single or double or treble or triple).mp. (42388) 19 (mask$ or blind$ or cross over or crossover or follow up).mp. (18036) 20 18 and 19 (3663) 21 4 or 5 or 6 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 20 (24589) 22 3 and 21 (9) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (84) 24 (glatiramer acetate or Copaxone).mp. (26) 25 22 and 23 (3) 26 limit 25 to english language (3) 27 limit 26 to yr="2001 - 2007" (3) 28 22 and 24 (2) 29 limit 28 to (english language and yr="2004 - 2007") (2) 30 (comment or letter or editorial or review).pt. (26645) 31 27 not 30 (3) 32 29 not 30 (2) 33 from 31 keep 1-3 (3) 34 from 32 keep 1-2 (2)	
EMBASE <1996 to 2006 Week 50>	1 random$.ti,ab. (221123) 2 factorial$.ti,ab. (4610) 3 (crossover$ or crossover$ or cross-over$).ti,ab. (18573) 4 placebo$.ti,ab. (55546) 5 (doubl$ adj blind$).ti,ab. (39320) 6 (singl$ adj blind$).ti,ab. (3538) 7 Crossover Procedure/ (14161) 8 Double Blind Procedure/ (43369) 9 Randomized Controlled Trial/ (98661) 10 Single Blind Procedure/ (5330) 11 or/1-10 (275641) 12 exp ANIMAL/ or NON HUMAN/ or exp ANIMAL EXPERIMENT/ (471846) 13 exp HUMAN/ (3120290) 14 12 and 13 (47560) 15 12 not 14 (424286) 16 11 not 15 (256075) 17 Multiple Sclerosis/ (17742) 18 (((Multiple Sclerosis or MS) adj5 Relapsing Remitting) or RR?MS). 1696	29

	19 17 or 18 (17891) 20 *beta interferon/ or *beta1 interferon/ (1950) 21 beta interferon/ or beta1 interferon/ (5215) 22 (letter or editorial or "review").pt. (916559) 23 (interferon beta or interferon-beta or Avonex or Rebif or Beta?eron).mp. (3220) 24 21 or 23 (7081) 25 16 and 19 and 24 (358) 26 limit 25 to (english language and yr="2001 - 2007") (232) 27 26 not 22 (166) 28 (glatiramer acetate or Copaxone).mp. (755) 29 Glatiramer/ (1263) 30 28 or 29 (1340) 31 16 and 19 and 30 (152) 32 limit 31 to (english language and yr="2004 - 2007") (51) 33 32 not 22 (29) 34 from 27 keep 1-166 (166) 35 from 33 keep 1-29 (29)	
	TOTAL refs in database after deduplication Keyworded Include in User def 1 field (separate export)	45 8 Includes

7.3 Appendix 3 Quality assessment of included RCTs of natalizumab

Using NICE's quality assessment of RCTs

1. AFFIRM trial

Quality criteria	Description	
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	C
Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	B
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	B

Other questions:	Response
Was the design parallel-group or cross-over? Indicate for each cross-over trial whether a carry-over effect is likely.	Parallel group
Was the trial conducted in the UK (or were one or more centres of the multinational trial located in the UK)? If not, where was the trial conducted and is clinical practice likely to differ from UK practice?	Multicentre – 99 countries in Europe, North America and Australasia, including the UK. Not know how

	many people for the UK were included.
How do the subjects included in the trial compare with patients who are likely to receive the drug in the UK? Consider factors known to affect outcomes in the main indication such as demographics, epidemiology, disease severity, setting.	The trial population is those with RRMS – this may be a more moderately affected, younger population with fewer years since diagnosis than the RES and SOT populations. Although subgroup analysis is done for the RES group, this is post-hoc and reduces the power of the study. The study population appears to be relatively treatment naïve. It is possible that there are also differences in treatment approach and additional drug therapy by country.
For pharmaceuticals, what dosage regimens were used in the trial? Are they within those detailed in the Summary of Product Characteristics?	300mg IV infusion every 4 weeks
What was the median (and range) duration of follow-up in the trial?	91% completed the planned 120-week study. Median and range not given

2. SENTINEL trial

Quality criteria	**Description**	
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	C

Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	B
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	B

Other questions:	**Response**
Was the design parallel-group or cross-over? Indicate for each cross-over trial whether a carry-over effect is likely.	Parallel group
Was the trial conducted in the UK (or were one or more centres of the multinational trial located in the UK)? If not, where was the trial conducted and is clinical practice likely to differ from UK practice?	Multicentre – 124 centres in Europe and USA, including the UK. Not known if UK was included.
How do the subjects included in the trial compare with patients who are likely to receive the drug in the UK? Consider factors known to affect outcomes in the main indication such as demographics, epidemiology, disease severity, setting.	The trial population is those with RRMS – this may be a more moderately affected, younger population with fewer years since diagnosis than the RES, and SOT populations.
For pharmaceuticals, what dosage regimens were used in the trial? Are they within those detailed in the Summary of Product Characteristics?	300mg IV infusion every 4 weeks
What was the median (and range) duration of follow-up in the trial?	86% completed the planned 120-week study. Median and range not given. Study was stopped one month early due to safety alerts about PML – two cases of the three detected to date came from this study.

3. Miller and colleagues (2003, Study 231)[61]

Quality criteria	Description	
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	C
Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	B missing data was imputed. 6mg 7.7% 3mg 2.2% Placebo 8.2%
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	B

Other questions:	Response
Was the design parallel-group or cross-over? Indicate for each cross-over trial whether a carry-over effect is likely.	Parallel group
Was the trial conducted in the UK (or were one or more centres of the multinational trial located in the UK)? If not, where was the trial conducted and is clinical practice likely to differ from UK practice?	Multicentre – 26 centres in USA, Canada and the UK. Not known how many people for the UK were included.
How do the subjects included in the trial compare with patients who are likely to receive the drug in the UK?	The trial population is those with RRMS (n=52) or SPMS (n=22).

Consider factors known to affect outcomes in the main indication such as demographics, epidemiology, disease severity, setting.	natalizumab is not licensed for use in SPMS. The RRMS group may be a more moderately affected, younger population with fewer years since diagnosis than the RES and SOT populations. Although subgroup analysis is done for the RES group, this is post-hoc and reduces the power of the study. The study population appears to be relatively treatment naïve. It is possible that there are also differences in treatment approach and additional drug therapy by country.
For pharmaceuticals, what dosage regimens were used in the trial? Are they within those detailed in the Summary of Product Characteristics?	3mg/kg or 6mg/ kg of body weight IV infusion every 4 weeks
What was the median (and range) duration of follow-up in the trial?	95% completed the planned 6 month follow up. Median 3% and range 2-10% (highest in 6mg/kg group)

4. Trial MS201 (details taken from Appendix D of the submission – summary of MS201 and the critical appraisal in F.3 if this provided additional information.

Quality criteria	**Description**	
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	C

Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	Unclear
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	B

Other questions:	**Response**
Was the design parallel-group or cross-over? Indicate for each cross-over trial whether a carry-over effect is likely.	Parallel group
Was the trial conducted in the UK (or were one or more centres of the multinational trial located in the UK)? If not, where was the trial conducted and is clinical practice likely to differ from UK practice?	Multicentre with nine UK centres.
How do the subjects included in the trial compare with patients who are likely to receive the drug in the UK? Consider factors known to affect outcomes in the main indication such as demographics, epidemiology, disease severity, setting.	The trial population is those with RRMS (n=53) or SPMS (n=19). natalizumab is not licensed for use in SPMS. The RRMS group may be a more moderately affected, younger population with fewer years since diagnosis than the RES and SOT populations. Although subgroup analysis is done for the RES group, this is post-hoc and reduces the power of the study. The study population appears to be relatively treatment naïve. It is possible that there are also differences in treatment approach and additional drug therapy by country.

For pharmaceuticals, what dosage regimens were used in the trial? Are they within those detailed in the Summary of Product Characteristics?	3mg/kg of body weight IV infusion every 4 weeks
What was the median (and range) duration of follow-up in the trial?	95% completed the planned 6 month follow up. Median and range not given

7.4 Appendix 4: Quality assessment of systematic reviews of comparators

Assessed using the QUORUM checklist

Although the Biogen submission updates the searches for the systematic reviews of GA and INF – no additional studies were identified that met their inclusion criteria. Comparator data therefore comes form the original reviews alone. These are assessed below.

1. IFN in relapsing remitting multiple sclerosis (Review) Rice and colleagues, 2002

1. Title: Identify the report as a systematic review?

Yes – as a Cochrane review

2. Abstract: Uses a structured format?

Yes – organised as below

Background	Outlines the clinical problem.
Objectives	The clinical question states that the review will assess the impact of recombinant IFNs (ie including IFN-α and -β
Search strategy	Data bases and additional sources searched are listed.
Selection criteria	Describes the population, intervention including method of administration, and study design.
Data collection and analysis	Describes who undertook data extraction and quality assessment. Does not describes all outcomes extracted, methods of data synthesis or methods for validity assessment not described.
Main results	Amount of identified evidence described. Characteristics of included trials not reported. Description of findings presented including point estimates or CIs.
Reviewers' conclusions	Reports the main results qualitatively.

3. Introduction

Yes – Background section describes the clinical problem and biological rationale for the intervention.

4. Methods

Searching	Databases searched are listed, additional methods - hand searching and contact with researchers and manufacturers - also listed. No restrictions of publication status, language or year of publication are made. Study was conducted in 2000.
Selection	Inclusion criteria are given which include description of included population, intervention study design and outcomes.
Validity assessment	Methodological quality is described in relation to adequate blinding, definition and relevance of clinical outcomes, number and handling of withdrawals and ITT analysis. Where data was unclear, trial sponsors were contacted for information. One study was excluded because neither blinding criteria nor clinical outcomes were clearly described.
Data extraction	Independently by four reviewers.
Study characteristics	Study design, patient characteristics, intervention details, and outcome definitions assessed. Heterogeneity assessed using chi-square tests.
Quantitative data synthesis	Dichotomous data expressed as Relative risks and 95% CI, meta-analysis using RevMan, continuous data shown as weighted mean difference and 95% CI. Fixed effects approach used unless significant heterogeneity – 9 pooled outcomes used random effects and 8 used fixed effects models.

5. Results

Trial flow	No diagram but number of studies identified and included described, together with reasons for exclusion. Seven trials (6 in IFN-β B and 1 in IFN-α) were included providing data on 1215 participants although key efficacy data was only available for 919.
Study characteristics	Study design, methods, patient characteristics, intervention details, outcome definitions, and Allocation concealment score are reported in the individual trial data extraction sheets, rather than being tabulated in the RevMan based paper., but are tabulated in the subsequent Lancet paper.
Quantitative data synthesis	Four reviewers extracted data but agreement on selection and validity assessment is not reported. Results of meta-analysis presented from RevMan.

	Compared to placebo in people with RRMS at two years: IFN reduces exacerbations RR 0.80 (95% CI 0.723 and , 0.88 p<0.001) IFN reduces disease progression RR 0.69 (95% CI 0.55, 0.87 p=0.02) More people receiving IFN experienced flu-like symptoms, fever, myalgia, fatigue, headache, injection site reactions, and hair loss.

6. Discussion

The discussion summarises key findings, clinical inferences based on internal and external validity are not discussed, the results are interpreted based on the total evidence included in the review, potential biases are not discussed. Sensitivity analyses was performed to assess the impact of the treatment of drop outs. A worst case scenario is to assume that all who dropped out progressed. In this analysis, the significance of the effect on disease progression is lost (RR 1.31, 95% CI 0.60, 2.89, p=0.5). Future research agenda is suggested.

Note that this systematic review was subject to several letter of criticism after it was published in the Lancet (see 2003; 361 p 1821-24). In particular they were criticised for combining trials of alpha- and beta- interferon, combining different doses of IFBN-B and assumptions made in speculative analyses about the status of drop-outs.

1. GA in relapsing remitting multiple sclerosis (Review) Munari and colleagues, 2003

7. Title: Identify the report as a systematic review?

Yes – as a Cochrane review

8. Abstract: Uses a structured format?

Yes – organised as below

Background	Outlines the clinical problem.
Objectives	To perform a Cochrane review of all placebo controlled trials of GA in MS, whatever the disease course.
Search strategy	Data bases and additional sources searched are listed.
Selection criteria	Describes the population, intervention and study design.
Data collection and analysis	Lists populations included, and quality included studies in.

	Does not describe who undertook data extraction and quality assessment, outcomes extracted, methods of data synthesis or methods for validity assessment.
Main results	Amount of identified evidence described. Characteristics of included trials not reported. Description of findings presented including point estimates or CIs.
Reviewers' conclusions	Reports the main results qualitatively.

9. Introduction

Yes – Background section describes the clinical problem and biological rationale for the intervention.

10. Methods

Searching	Databases searched are listed, additional methods - hand searching and contact with researchers and manufacturers - also listed. No restrictions of publication status, language or year of publication are made. However, while publication date is said to be 2003, the dates for the searches are up to 2004.
Selection	Inclusion criteria are given which include description of included population, intervention study design and outcomes.
Validity assessment	Methodological quality is assessed by a Jadad score.
Data extraction	Independently by 2 reviewers. Trialists contacted where information was absent or unclear.
Study characteristics	Study design, patient characteristics, intervention details, and outcome definitions assessed. Heterogeneity assessed using chi-square tests.
Quantitative data synthesis	Dichotomous data expressed as relative risks and 95% CI, continuous outcomes used weighted mean difference in meta-analysis using RevMan, continuous data shown as weighted mean difference and 95% CI. Stated that a fixed effects approach was used unless significant heterogeneity – however all pooled outcomes used fixed effects models, despite demonstrating heterogeneity (for example, in the assessment of progression at 2 years, I^2 = 39%).

11. Results

Trial flow	No diagram but number of studies identified and included described, together with reasons for exclusion. Seventeen papers relating to four trials (three relating to RRMS and one relating to CPMS) were included providing data on 646 participants.
Study characteristics	Study design, methods, patient characteristics, intervention details, outcome definitions, and allocation concealment score are reported in individual trial data extraction sheets only, not tabulated.
Quantitative data synthesis	Two reviewers extracted data but agreement on selection and validity assessment is not reported. Results of meta-analysis presented from RevMan (excluding Bornstein CPMS pts). Compared to placebo in people with RRMS at two years: GA does not reduce the number of people having at least one exacerbation at 2 years RR 0.87 (95% CI 0.74, 1.02 p=0.08) GA does not reduce disease progression RR 0.77 (95% CI 0.51, 1.14 p=0.2) although a slight decrease in mean EDSS score at the end of study was seen (although note that this absolute difference is small and of unknown clinical meaning, and I2 in this fixed effects analysis = 61.1%) -0.33 (95% CI 0.58, 0.08) More people receiving GA experienced local site reactions, dizziness and palpitations.

12. Discussion

The discussion briefly summarises key findings. Clinical inferences based on internal and external validity are not discussed nor are potential review biases.

7.5 Appendix 5: Analysis of MS cohort, by treatment option, over time.

See Table 16 for the data extracted from the model on the proportion of the cohort (SOT and RES subgroup) in different model states over time, by treatment option.

Figure 4 a-d and Figure 5 a-d, show similar data graphically.

Table 18: PenTAG analysis of model data: (a) SOT, (b) RES subgroup cohort location over time, by treatment option (natalizumab, IFN-β and GA – BSC?).

(a)

	RRMS EDSS 0 - 3				RRMS EDSS 4 - 6				Proportion drop out				Proportion dead				Proportion treated			
Year	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**
1	0.78	0.78	0.78	0.78	0.22	0.22	0.22	0.22	n/a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	n/a	1.00	1.00	1.00
5	0.48	0.52	0.47	0.41	0.17	0.08	0.11	0.12	n/a	0.39	0.42	0.46	0.01	0.01	0.01	0.01	n/a	0.60	0.58	0.54
10	0.28	0.30	0.24	0.19	0.10	0.04	0.05	0.06	n/a	0.64	0.68	0.73	0.02	0.02	0.02	0.02	n/a	0.34	0.30	0.25
15	0.16	0.22	0.16	0.11	0.06	0.03	0.04	0.03	n/a	0.70	0.76	0.81	0.05	0.05	0.05	0.05	n/a	0.25	0.19	0.14
20	0.09	0.18	0.11	0.07	0.03	0.02	0.02	0.02	n/a	0.71	0.77	0.81	0.11	0.09	0.10	0.10	n/a	0.20	0.13	0.09
25	0.05	0.14	0.07	0.04	0.02	0.02	0.02	0.01	n/a	0.68	0.73	0.76	0.19	0.17	0.18	0.19	n/a	0.15	0.09	0.05
30	0.03	0.10	0.05	0.02	0.01	0.01	0.01	0.01	n/a	0.60	0.64	0.65	0.32	0.29	0.30	0.32	n/a	0.12	0.06	0.03

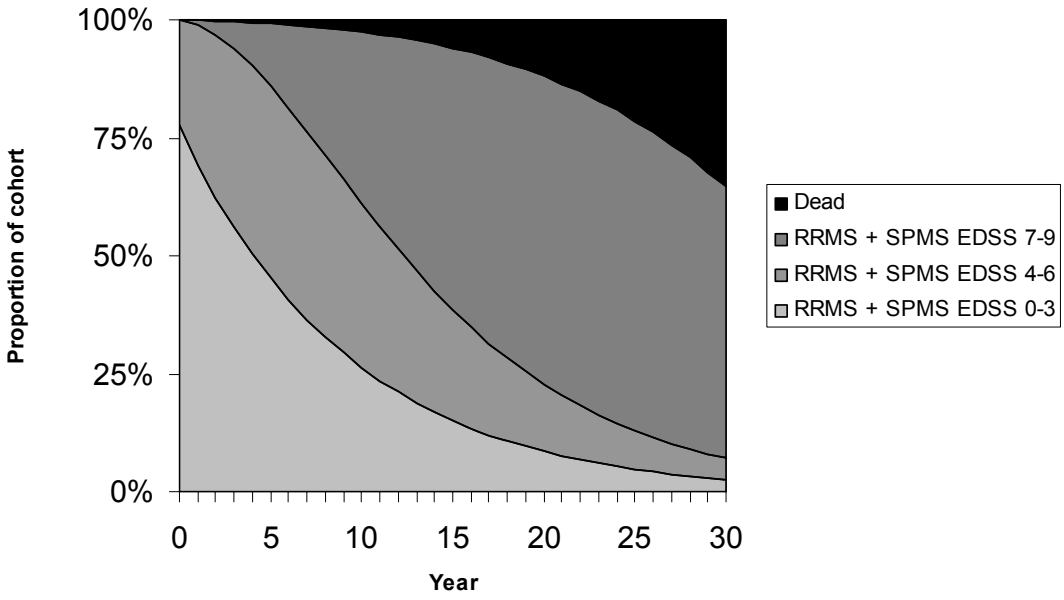
(b)

	RRMS EDSS 0 - 3				RRMS EDSS 4 - 6				Proportion drop out				Proportion dead				Proportion treated			
Year	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**	**BSC**	**NAT**	**IFN**	**GA**
1	0.78	0.78	0.78	0.78	0.22	0.22	0.22	0.22	n/a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	n/a	1.00	1.00	1.00
5	0.42	0.52	0.42	0.37	0.21	0.09	0.14	0.16	n/a	0.38	0.43	0.47	0.01	0.01	0.01	0.01	n/a	0.61	0.57	0.53
10	0.22	0.30	0.21	0.16	0.11	0.05	0.07	0.07	n/a	0.63	0.70	0.75	0.02	0.02	0.02	0.02	n/a	0.35	0.28	0.23
15	0.11	0.23	0.13	0.09	0.06	0.04	0.04	0.04	n/a	0.69	0.78	0.83	0.05	0.05	0.05	0.05	n/a	0.26	0.17	0.12
20	0.06	0.18	0.08	0.05	0.03	0.03	0.03	0.02	n/a	0.70	0.79	0.83	0.11	0.09	0.10	0.11	n/a	0.21	0.11	0.07
25	0.03	0.14	0.05	0.03	0.01	0.02	0.02	0.01	n/a	0.67	0.75	0.77	0.20	0.17	0.18	0.19	n/a	0.16	0.07	0.04
30	0.01	0.10	0.03	0.01	0.01	0.02	0.01	0.01	n/a	0.59	0.64	0.65	0.33	0.29	0.31	0.32	n/a	0.12	0.04	0.02

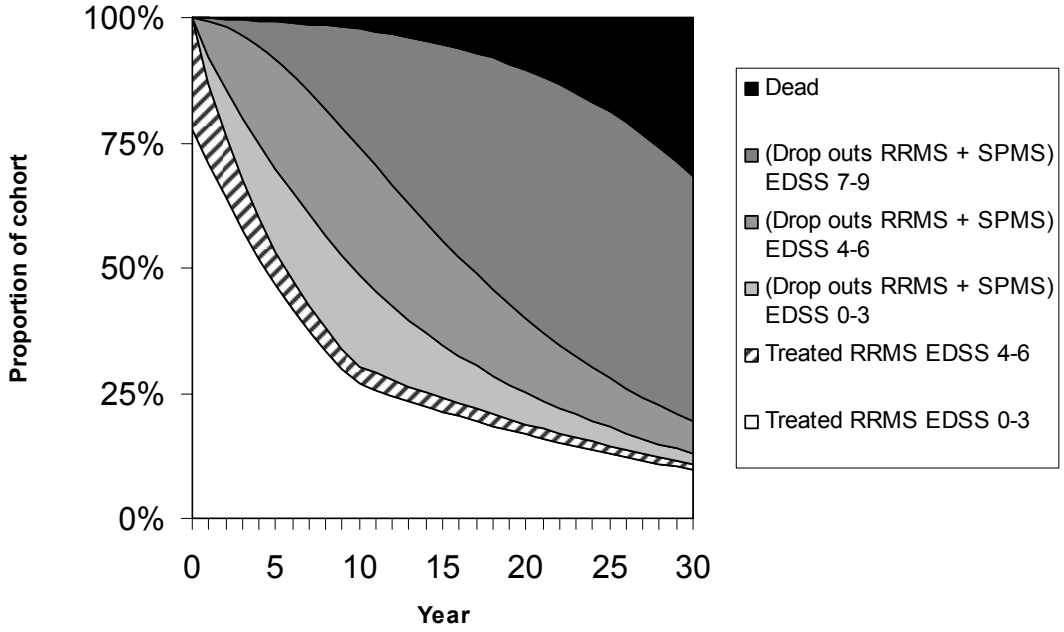
NB. The proportion of patients in RRMS EDSS 7-9 is very low at all time points, at most 1%.

Figure 4: SOT subgroup cohort split over time for (a) placebo, (b) NAT, (c) IFN-β, (d) GA.

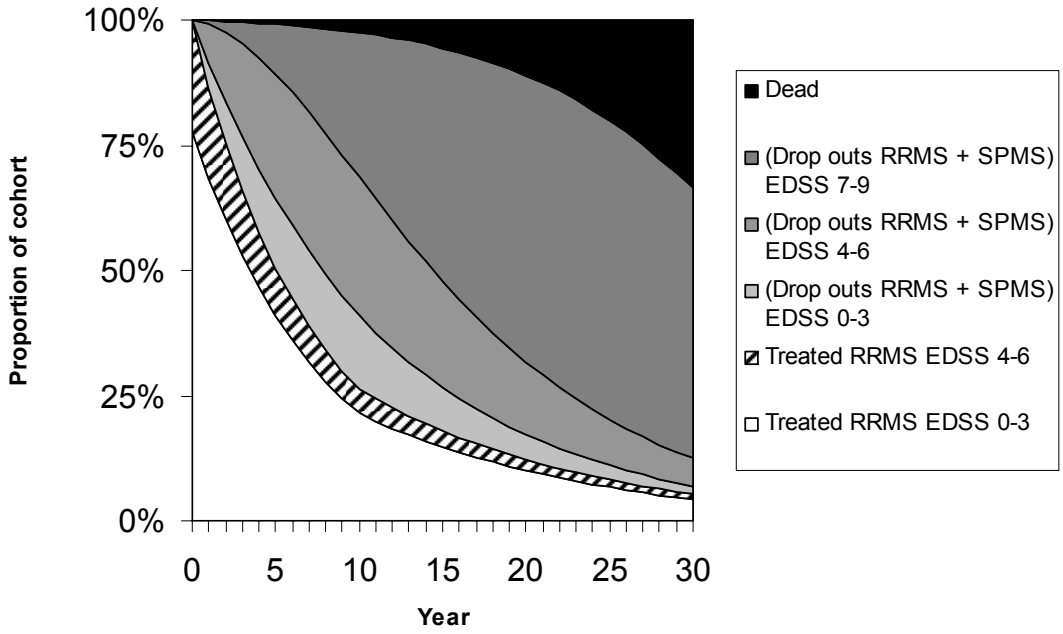
(a) SOT Placebo



(b) SOT natalizumab



(c) SOT Beat-interferon



(d) SOT Glatiramer acetate

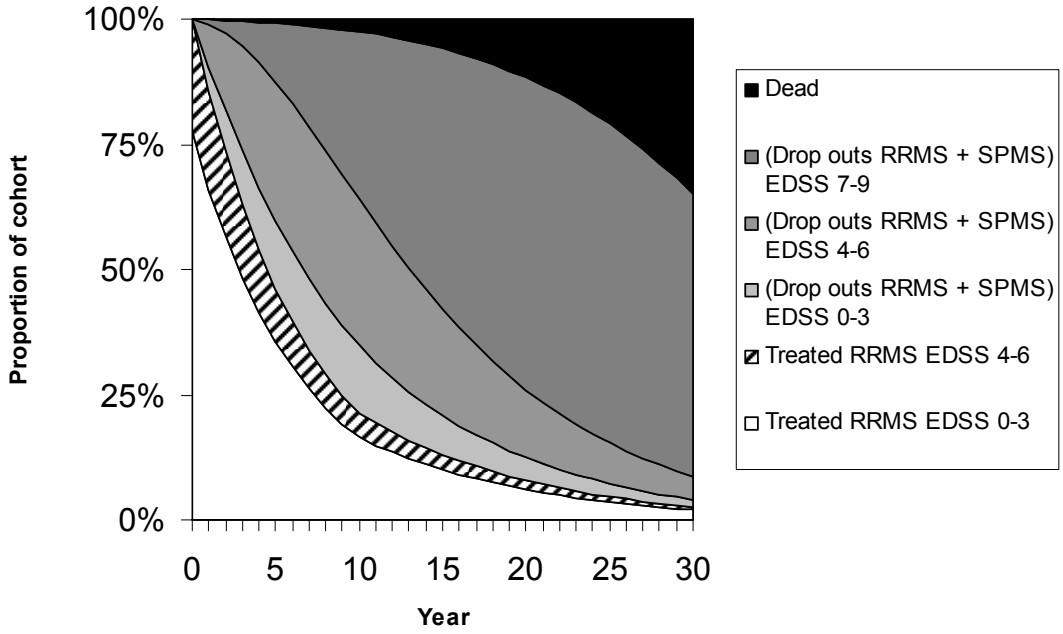
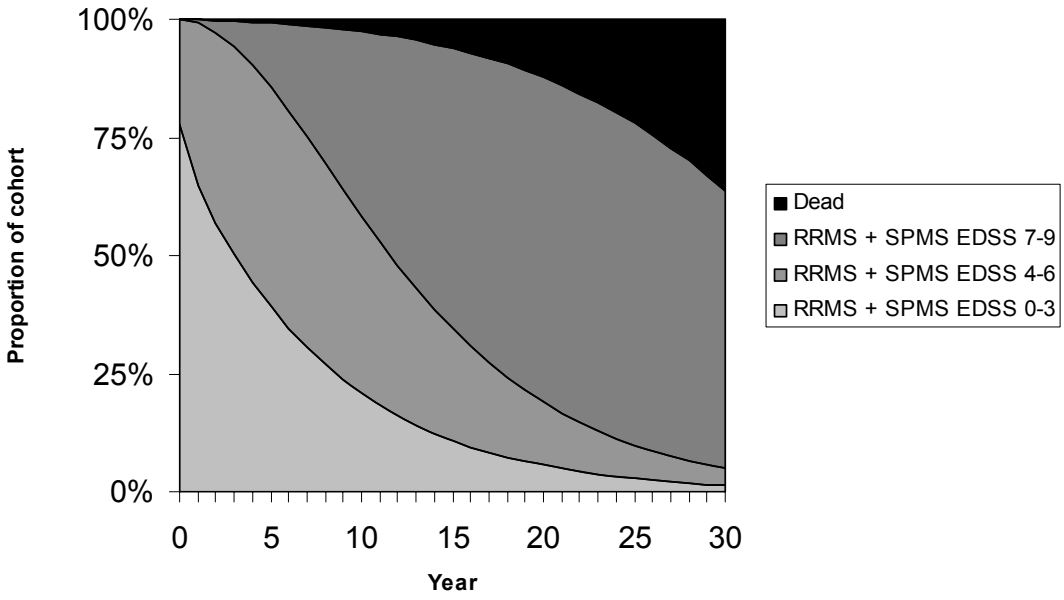
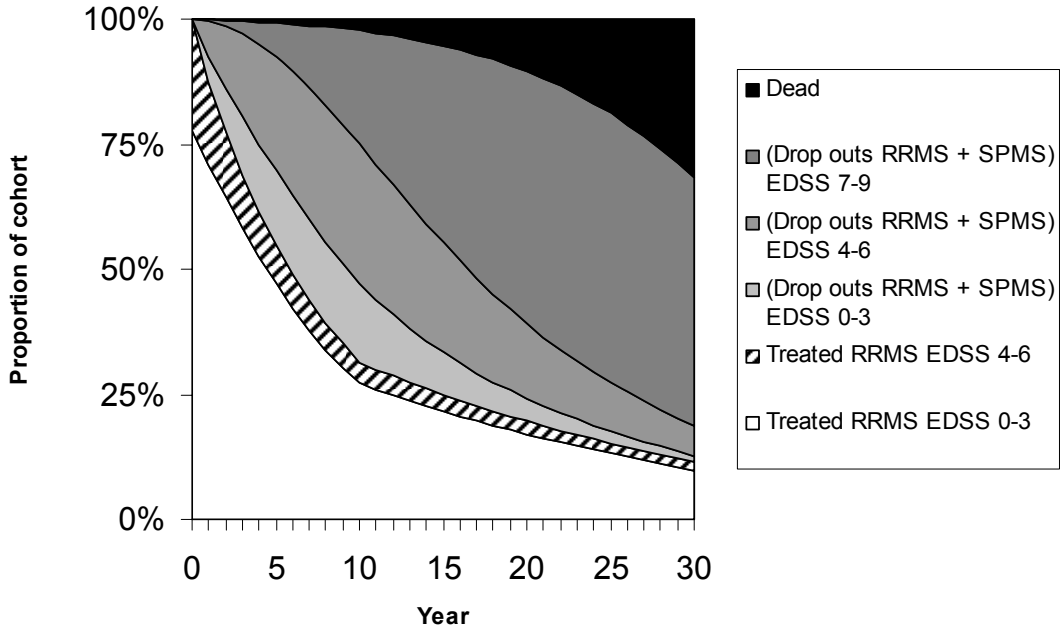


Figure 5: RES subgroup cohort split over time for (a) placebo, (b) natalizumab, (c) IFN-β, (d) GA.

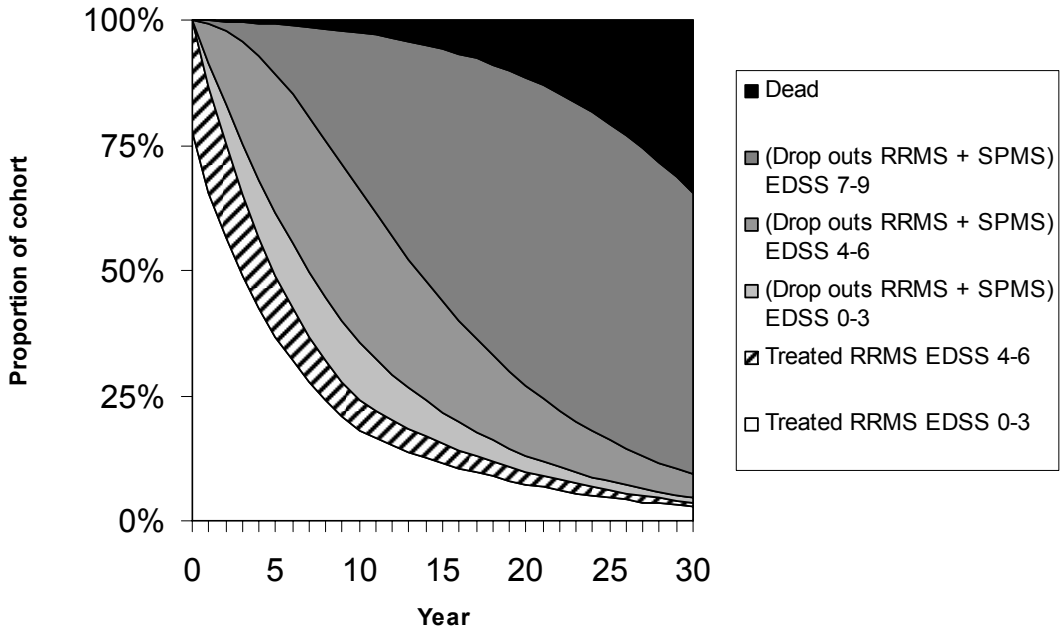
(a) RES Placebo



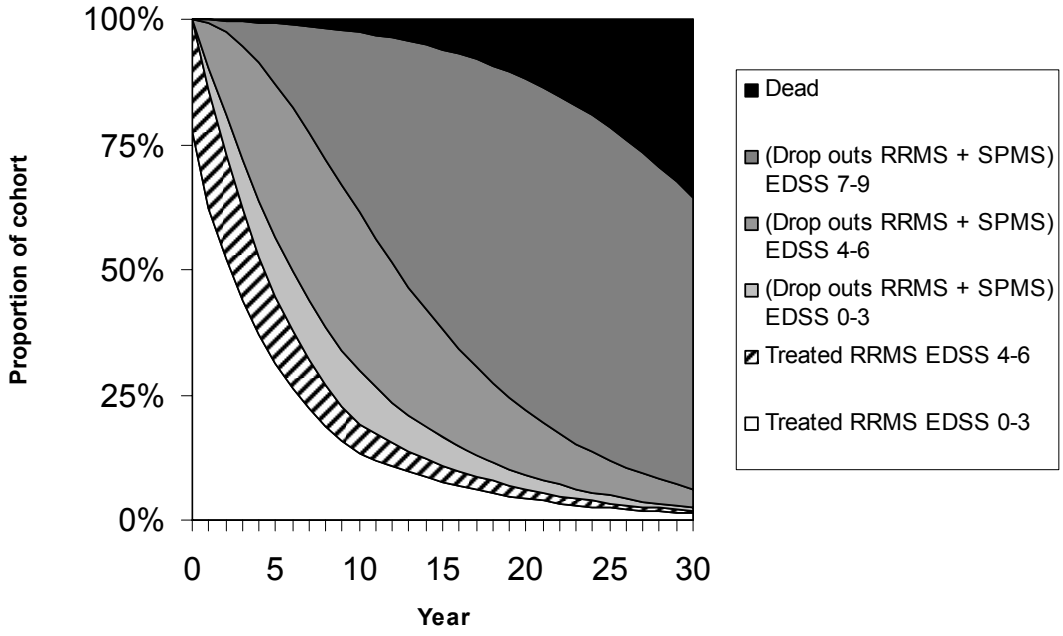
(b) RES natalizumab



(c) RES Beta-interferon



(d) RES Glatiramer acetate



7.6 Appendix 6: Importance of relapse rate in cost-effectiveness of natalizumab.

Table 19: Effect of NAT relapse rate on ICERs

Patient subgroup	ICER under base case (NAT relapse and progression rates unchanged)	ICER change (,000s) given NAT relapse rate equal to comparator
SOT NAT vs. BSC	56,138	1.8
SOT vs. IFN-β	43,414	1.4
SOT NAT vs. GA	44,341	1.2
RES NAT vs. BSC	44,604	3.1
RES NAT vs. IFN-β	32,027	2.0
RES NAT vs. GA	34,559	1.9

In a separate analysis, we found that applying the same disability progression data for all comparators and only using differential relapse rates led to very high cost per QALY estimates.

7.7 Appendix 7: EDSS Health State Costs

Table 20: EDSS specific costs reported as an addendum by Tappenden and colleagues 2001

	RRMS	SPMS
EDSS	**Cost (£s)**	**Cost (£s)**
0	756	756
1.0	756	756
1.5	756	756
2.0	756	756
2.5	1,394	1,394
3.00	1,394	1,394
3.5	1,444	1,444
4.0	1,444	1,444
4.5	5,090	5,090
5.0	5,090	5,090
5.5	5,678	5,678
6.0	5,678	5,678
6.5	11,445	11,445
7.0	17,327	17,327
7.5	17,327	17,327
8.0	26,903	26,903
8.5	26,903	26,903
9.0	34,201	34,201
9.5	34,201	34,201
10	0	0
Relapse	2,697	2,697

Note: Data source and methods used to derive estimates are unknown.

Reference List

1. The National Collaborating Centre for Chronic Conditions. Multiple Sclerosis. National clinical guideline for diagnosis and management in primary and secondary care. 2004. London, Royal College of Physicians.

Ref Type: Report

2. Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue EW *et al*. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:911-23.

3. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

4. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

5. National Institute for Clinical Excellence. Beta-interferon and glatiramer acetate for the treatment of multiple sclerosis. 2002. London, National Institute for Clinical Excellence. Technology Appraisal Guidance - No. 32.

Ref Type: Report

6. Rice G, Incorvaia B, Munari L, Ebers GC, Polman C, D'Amico R *et al*. Interferon in relapsing-remitting multiple sclerosis. *Cochrane Database of Systematic Reviews* 2001;**4**.

7. Munari L, Lovarti R, Boiko A. Therapy with glatiramer acetate for multiple sclerosis. *Cochrane Database of Systematic Reviews* 2003.

8. Martinelli Boneschi F, Rovaris M, Capra R, Comi G. Mitoxantrone for Multiple Sclerosis. *Cochrane Database of Systematic Reviews* 2005.

9. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

10. Prognostic indicators of future disease activity in patients with relapsing multiple sclerosis (MS): an analysis from the AFFIRM study. *Journal of the Neurological Sciences* 2005;**238**.

11. The effects of baseline disease activity on magnetic resonance imaging (MRI) outcomes in the AFFIRM study. *Journal of the Neurological Sciences* 2005;**238**:S75-S76.

12. Fisher E, O'Connor R, Havrdova E, Hutchinson M, Kappos L, Miller D *et al*. The effect of natalizumab on brain atrophy in patients with relapsing multiple sclerosis: 2-year results from the AFFIRM study. *Journal of Neurology* 2005;**252**:149.

13. Havrdova E, O'Connor P, Hutchinson M, Kappos L, Miller D, Phillips JT *et al*. The efficacy of natalizumab in patients with relapsing multiple sclerosis: subgroup analyses of AFFIRM. *Journal of Neurology* 2005;**252**:46.

14. Lublin FD, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. The effects of natalizumab on disability progression in the AFFIRM study: correlation between changes in Multiple Sclerosis Functional Composite and Expanded Disability Status Scale scores. *Multiple Sclerosis* 2005;**11**:S152-S153.

15. Miller D, O'Connor P, Havrdova E, Hutchinson M, Kappos L, Phillips JT *et al*. Magnetic resonance imaging results from AFFIRM: a randomised controlled trial of natalizumab in patients with relapsing multiple sclerosis. *Journal of Neurology* 2005;**252**:46-7.

16. Miller D, O'Connor P, Havrdova E, Hutchinson M, Kappos L, Phillips T *et al*. The efficacy of natalizumab on magnetic resonance imaging (MRI) measures in patients with relapsing multiple sclerosis (MS): Results from the AFFIRM trial. *Neurology* 2005;**64**:A147.

17. O'Connor P, Havrdova E, Hutchinson M, Kappos L, Phillips J, Polman C *et al*. Safety, tolerability and immunogenicity of natalizumab: Results from the AFFIRM trial. *Neurology* 2005;**64**.

18. Miller DH, Khan OA, Sheremata WA, Blumhardt LD, Rice GP, Libonati MA *et al*. A controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2003;**348**:15-23.

19. O'Connor PW, Goodman A, Willmer-Hulme AJ, Libonati MA, Metz L, Murray RS *et al*. Randomized multicenter trial of natalizumab in acute MS relapses: clinical and MRI effects.[see comment]. *Neurology* 2004;**62**:2038-43.

20. Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue EW *et al*. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:911-23.

21. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

22. Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue EW *et al*. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:911-23.

23. Miller DH, Khan OA, Sheremata WA, Blumhardt LD, Rice GP, Libonati MA *et al*. A controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2003;**348**:15-23.

24. Chofflon M. Mechanisms of action for treatments in multiple sclerosis does a heterogeneous disease demand a multi-targeted therapeutic approach? *Biodrugs* 2005;**19**:299-308.

25. Goodman A, Bar-Or A, Miller A, Miller D, Panzara M, Rossman H *et al*. GLANCE: A double-blind, randomised, placebo-controlled, parallel-group safety study of natalizumab (Antegren) in combination with glatiramer acetate (Copaxone) in subjects with relapsing-remitting multiple sclerosis. *Journal of Neurology* 2004;**251**:176-7.

26. European Medicines Agency. Guideline on Clinical Investigation of Medicinal Products for the Treatment of Multiple Sclerosis. 16-12-2006. London, Eurpoean Medicines Agency.
Ref Type: Report

27. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

28. European Medicines Agency. Guideline on Clinical Investigation of Medicinal Products for the Treatment of Multiple Sclerosis. 16-12-2006. London, Eurpoean Medicines Agency.
Ref Type: Report

29. European Medicines Agency. Guideline on Clinical Investigation of Medicinal Products for the Treatment of Multiple Sclerosis. 16-12-2006. London, Eurpoean Medicines Agency.
Ref Type: Report

30. Hobart JC, Freeman J, Thompson AJ. Kurtzke scales revisted: the application of psychometric methods to clincial intuition. *Brain* 2000;**123**:1027-40.

31. Thompson AJ,.Hobart JC. Multiple sclerosis: assessment of disability and disability scales. *Journal of Neurology* 1998;**245**:189-96.

32. European Medicines Agency. Guideline on Clinical Investigation of Medicinal Products for the Treatment of Multiple Sclerosis. 16-12-2006. London, Eurpoean Medicines Agency.
Ref Type: Report

33. Langer-Gould A,.Steinman L. What went wrong in the natalizumab trials? *Lancet* 2006;**367**:708-10.

34. Scottish Medicines Consortium. Natalizumab 300mg concentrate for solution for infusion. 329/06. 10-11-2006. NHS Scotland.
Ref Type: Report

35. Phillips, J. T., Basel, L K., O'Connor, P. W., Havrdova, E., Hutchinson, M., Miller, D. H., Polman, C., Lublin, F., Panzara, M., and Sandrock, A. W. The Effects of Natalizumab Monotherapy on Multiple Measures of Disability Progression in MS. 58th Annual Meeting of the American Academy of Neurology. Multiple Sclerosis: Clincial Trials . 4-4-2006. San Diego, California. 2006.
Ref Type: Conference Proceeding

36. Chaudhuri A. Lessons for clinical trials from natalizumab in multiple sclerosis. [Review] [20 refs]. *BMJ* 2006;**332**:416-9.

37. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J.Clin.Epidemiol.* 1997;**50**:683-91.

38. Chaudhuri A. Lessons for clinical trials from natalizumab in multiple sclerosis. [Review] [20 refs]. *BMJ* 2006;**332**:416-9.

39. Chaudhuri A. Lessons for clinical trials from natalizumab in multiple sclerosis. [Review] [20 refs]. *BMJ* 2006;**332**:416-9.

40. Chaudhuri A. Lessons for clinical trials from natalizumab in multiple sclerosis. [Review] [20 refs]. *BMJ* 2006;**332**:416-9.

41. Langer-Gould A,.Steinman L. What went wrong in the natalizumab trials? *Lancet* 2006;**367**:708-10.

42. Yousry TA, Major EO, Ryschkewitsch C, Fahle G, Fischer S, Hou J *et al*. Evaluation of patients treated with natalizumab for progressive multifocal leukoencephalopathy.[see comment]. *New England Journal of Medicine* 2006;**354**:924-33.

43. Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1997.

44. Phillips CJ. The cost of multiple sclerosis and the cost effectiveness of disease-modifying agents in its treatment. [Review] [98 refs]. *CNS Drugs* 2004;**18**:561-74.

45. Hobart JC, Kalkers J, Barkhof F, Uitdehaag BM, Polman C, Thompson AJ. Outcome measures for multiple sclerosis clinical trials: relative measurement precision of the Expanded Disability Status Scale and Multiple Sclerosis Functional Composite. *Multiple Sclerosis* 2004;**10**:41-6.

46. Patzold U,.Pocklington PR. Course of Multiple Sclerosis. First results pof a prospective study cvarried out of 102 MS patients from 1976-1980. *Acta Neurol Scandinav* 1982;**65**:248-66.

47. Pokorski RJ. Long-term survival experience of patients with multiple sclerosis. *Journal of Insurance Medicine (Seattle)* 1997;**29**:101-6.

48. Tappenden, P., Chilcott, J., O'Hagan, A., McCabe, C., Cooper, N. J., Abrams, K., Ebers, G. C., Fenwick, L., McNamee, P., Parkin, D., and Sculpher, M. Cost-effectiveness of beta interferons and glatimer acetate in the mangement of multiple sclerosis. Final report to NICE. 2001. School of Health and Related Research, Universoty of Sheffield.
Ref Type: Report

49. Herndon RM, Rudick RA, Munschauer FE, III, Mass MK, Salazar AM, Coats ME, and et al. Eight-year immunogenicity and safety of interferon beta-1a-Avonex

treatment in patients with multiple sclerosis. Multiple Sclerosis 11(4), 409-419. 2005.
Ref Type: Journal (Full)

50. Kobelt G, Berg J, Lindgren P, Fredrikson S, Jonsson B. Costs and quality of life of patients with multiple sclerosis in Europe. *Journal of Neurology, Neurosurgery & Psychiatry* 2006;**77**:918-26.

51. Orme M, Kerrigan J, Tyas D, Russell N, and Nixon R. The effect of disease, disability and relapses on the utility of people with multiple sclerosis in the UK. Value Health in press. 2007.
Ref Type: Journal (Full)

52. Dolan P and Gudex C. A Social tariff for EuroQol: Results from a UK General Population Survey. Centre for Health Economics. 138. 1995. University of York.
Ref Type: Report

53. Parkin D, McNamee P, Jacoby A, and et al. A cost utility analysis of interferon beta for multiple sclerosis. Health Technology Assessment 2, iii-54. 1998.
Ref Type: Journal (Full)

54. Forbes RB, Lees A, Waugh N, and Swingler RJ. Population based cost utility study of interferon beta-1b in secondary progressive multiple sclerosis. BMJ 319, 1529-1533. 1999.
Ref Type: Journal (Full)

55. Gruenewald DA, Higginson IJ, Vivat B, and et al. Quality of life measures for the palliative care of people severely affected by multiple sclerosis: a systematic review. Multiple Sclerosis 10, 690-704. 2004.
Ref Type: Journal (Full)

56. Mitchell AJ, Benito-Leon J, Gonzalez JM, and Rivera-Navarro J. Quality of life and its assessment in multiple sclerosis: integrating physical and psychological components of wellbeing. Lancet Neurology 4, 556-566. 2005.
Ref Type: Journal (Full)

57. Prosser LA, Kuntz KM, Bar-Or A, and Weinstein MC. Patient and community preferences for treatments and health states in multiple sclerosis. Multiple Sclerosis 9(3), 311-319. 2003.
Ref Type: Journal (Full)

58. Tyas D, Kerrigan J, Russell N, and Nixon R. The distribution of the cost of multiple sclerosis in the UK: How do costs vary by illness severity? 2006.
Ref Type: Report

59. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

60. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2006;**354**:899-910.

61. Miller DH, Khan OA, Sheremata WA, Blumhardt LD, Rice GP, Libonati MA *et al*. A controlled trial of natalizumab for relapsing multiple sclerosis.[see comment]. *New England Journal of Medicine* 2003;**348**:15-23.