Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

IMI PROTECT WP5 IMI Report 2:b:iv

NATALIZUMAB WAVE 2 CASE STUDY REPORT

Review of methodologies for benefit and risk assessment of medication.

Risk benefit case study with a focus on testing methodology.

Four of four: natalizumab

Richard Nixon, Ed Waddingham, Shahrul Mt-Isa, Kimberley Hockley, Adam Elmachtoub, Dave Gelb, Rick Hermann, Silvia Kuhls, Alain Micaleff, Isabelle Stoeckert , Andrew Thomson, Nan Wang

On behalf of PROTECT Work Package 5 participants

Version 1 Date: 31 Mar 2013 Date of any subsequent amendments below	Person making amendments	Brief description of amendments

Acknowledgements: The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency.
The PROTECT project has received support from the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement n° 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution

Contents

Glossary

AE	Adverse Effect
BRA	Benefit-Risk Analysis
BRAT	Benefit-Risk Action Team
CHMP	Committee for Medicinal Products for Human Use
CIS	Clinically Isolated Syndrome
CNS	Central Nervous System
DALY	Disability Adjusted Life Year
DMD	Disease-Modifyng Drug
DSA	Deterministic Sensitivity Analysis
EDSS	Expanded Disability Status Scale
EMA	European Medicines Agency
EMSP	European Multiple-Sclerosis Platform
EPAR	European Public Assessment Report
EU	European
FDA	Food and Drug Administration
GA	Glatiramer Acetate
Gd	Gadolinium
HR	Hazard Ratio
IFN	Interferon
IM	Intramuscular
IMI	Innovative Medicines Initiative
ITT	Intention to Treat
IV	Intravenous
LCI	Lower Confidence Interval
MCDA	Multiple Criteria Decision Analysis
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSFC	Multiple Sclerosis Functional Composite
NCB	Net Clinical Benefit
NNH	Number Needed To Harm
NNT	Number Needed To Treat

Pbo	Placebo
PD	Pharmacodynamic
PK	Pharmacokinetic
PML	Progressive Multifocal Leukoencephalopathy
PPMS	Primary Progressive Multiple Sclerosis
PSA	Probabilistic Sensitivity Analysis
QALY	Quality Adjusted Life Year
RR	Relative Risk
RRMS	Relapsing-Remitting Multiple Sclerosis
SC	Subcutaneous
SPMS	Secondary Progressive Multiple Sclerosis
Tx	Active treatment
UCI	Upper Confidence Interval
UK	United Kingdom
US	United States
VAS	Visual Analogue Scale
w-NCB	Weighted Net Clinical Benefit

Section 1 Introduction and background

Part 1.1 Organisation of the report

This document details the second wave of a case study for performing a benefit-risk assessment of natalizumab as part of the IMI PROTECT Work Package 5.

Section 1 sets out the background to the case study, including general information on natalizumab and a summary of the wave 1 case study. The aims of this extended wave 2 case study are defined in section 2.

The work on visual representations and probabilistic uncertainty are reported in detail in Sections 3 and 4 respectively. Section 5 contains some discussion of the entire project and the conclusions drawn from the case study. Supporting information is provided in the Appendix.

Part 1.2 Multiple Sclerosis Disease Background

1.2.1 Concept and Pathology

Multiple Sclerosis (MS) is the most common neurological disorder causing disability in young adults affecting approx. 1 in 1.000 people in western countries. MS begins at the age of 20 to 40 years with a median age of 28 years at onset with acute episodes of neurological dysfunction, followed by periods of partial or complete remission and clinical stability in between relapses (RRMS) (1). MS is an autoimmune disease that occurs in genetically susceptible individuals exposed to specific environmental triggers. Apart from demyelination by autoimmune disturbance, the other pathologic hallmark of the disease is inflammation. According to a strictly immunological explanation of MS, the inflammatory process is caused by T cells. Those trigger inflammatory processes, stimulating other immune cells and soluble factors like cytokines and antibodies. Leaks form in the blood–brain barrier, which in turn cause a number of other damaging effects such as swelling, activation of macrophages, and more activation of cytokines and other destructive proteins. Approximately 85% of patients suffer from relapsing-remitting MS. However, this descriptive term is misleading because ultimately signs and symptoms of MS tend to progress, highlighting the need for prompt and consistent therapy. Early treatment has been advocated, including for clinically isolated syndrome (CIS), often a precursor to MS (2).

1.2.2 Diagnosis and Forms of Multiple Sclerosis

A guideline from the regulators perspective provides a clear definition of the indication under discussion (3). The term relapsing MS includes 1) patients with RRMS, 2) patients with Secondary progressive multiple sclerosis (SPMS) and superimposed relapses and 3) patients with a single demyelinating clinical event who show lesion dissemination on subsequent MRI scans according to McDonald's criteria. Prevention and/or modification of relapse features as well as prevention or delay of the accumulation of disability as sequela of acute relapses, are meaningful goals in the treatment of RMS. Progression of disability, as a result of relapses from which patients do not fully recover take many years and, for the moment, there are no surrogate variables for evaluating progression of disability.

Therefore large-scale long-term parallel group trials are required to establish clinically relevant treatment differences on disease progression. Relapse rate, relapse duration, and recovery after relapses are all highly variable between patients and for individual patients. Therefore, treatments intended to decrease the relapse rate or modify relapses should be evaluated in parallel trials sufficiently large and long to overcome this inter and intra-individual variability.

Currently approved therapies have demonstrated a favourable effect on the rate and severity of relapses with some products similar in the short-term (a few years) progression of disability. If a product demonstrates a benefit in relapse rate or severity without an accompanying effect on preventing or delaying disability, the clinical relevance of

such benefit should be justified. It is therefore accepted that the indication in relapsing MS will mainly rely on the effects shown in patients with RRMS, but may be extrapolated to an effect on relapses in SPMS.

Several major placebo-controlled clinical trials have provided evidence of an apparent short-term stabilisation in placebo-treated patients that could be explained, among others, by the regression to the mean phenomenon, and by a real placebo effect, as well as by the natural course of the disease. Approved therapies have been shown to favourably modify the short-term evolution of the disease although the benefit is modest, at the cost of significant patient inconveniences and side effects. Differences from placebo seem not always consistent across trials and the sensitivity of the available scales to measure progression of disability as well as other characteristics of clinical trials in this field do not assure the ability to detect clinically relevant differences.

1.2.3 Medical need and available treatment options

In MS, disability accrues during both acute inflammatory events and progressive phases of disease. In short to intermediate-term studies, suppression of acute inflammatory events reduced the risk of disability progression in MS. Therefore, early recognition of sub-optimal response to disease-modifying therapy (DMT) and prompt intervention may limit future impairment. Interventions for the treatment of MS include immunomodulatory agents such as interferon-beta or glatiramer acetate, monoclonal antibodies such as natalizumab or immunsuppressants such as mitoxantrone, azathioprine or cyclophosphamide. Very recently, fingolimod was approved and added to the list of disease-modifying agents. All of these agents reduce the frequency and severity of relapses to varying degrees, delay disability progression and decrease brain lesion development as evidenced by MRI. The Multiple Sclerosis Consensus Therapy Group (MSTCG) scheme considers interferon beta and glatiramer acetate as basic therapies whereas natalizumab and mitoxantrone represent possibilities for escalating treatment.

Natalizumab (Tysabri®) is a monoclonal antibody against α4-integrin, a component of VLA-4 located on leukocytes. Binding to VLA-4 blocks its interaction with the ligand VCAM (vascular cell adhesion molecule) on the surface of endothelial cells at the blood-brain barrier, thus greatly reducing the transmigration of lymphocytes and monocytes from venules and capillaries into inflamed tissue. In EU, natalizumab is indicated as a “disease-modifying monotherapy of highly active relapsing MS” for the following patient groups: 1) patients showing high levels of disease activity despite treatment with an IFN-β preparation, or 2) untreated/treatment-naive patients with rapidly progressing relapsing-remitting MS (at least two serious relapses per year).

1.2.4 Natalizumab and Progressive multifocal leukoencephalopathy

Progressive multifocal leukoencephalopathy (PML) is a brain infection caused by activation of the John Cunningham Virus (JCV), which is a type of human polyomavirus that is commonly found in the general population, but only leads to PML if the immune system has been compromised. Progressive multifocal leukoencephalopathy is a clinical manifestation of direct and active JCV infection of the oligodendrocytes, which leads to decreased myelin production and subsequent demyelination with resulting severe disability and ultimately, death. The signs and symptoms of PML can initially be clinically very similar to those of an MS attack due to the demyelination.

In October 2009 the CHMP together with the European Commission initiated an Article 20 review of the benefits and risks for natalizumab in view of the cases of PML that have been observed since natalizumab has been on the market and in consideration of the occurrence of Immune Reconstitution Inflammatory Syndrome (IRIS) in these patients once natalizumab has been stopped and Plasma Exchange (PLEX) and/or immunoabsorption (IA) was implemented.

The discussion on the validity as confirmatory or not of a positive CSF PCR for JCV was also discussed at the CHMP. The conclusion was that a negative CSF PCR for JCV, due to a potentially low number of viral copies, does not automatically mean the exclusion of PML.

1.2.4.1 PML data and discussion at time of initial EU approval

At the time of initial approval in 2006, two cases of PML, one fatal, had been reported by the Applicant in two patients with RRMS, both treated with a combination of natalizumab (Tysabri) and beta-interferon (Avonex) for more than 2 years. A third case of PML was later discovered upon re-evaluation of the Crohn's Disease safety database in a subject that was originally presumed to have died of a malignant astrocytoma. No confirmed cases of PML in the mono-therapy MS clinical trials were identified in the post-marketing setting. Over 4,500 person-years of experience with natalizumab exist between the MS and Crohn's Disease programmes. Of the 1,617 MS subjects treated with natalizumab in the placebo-controlled experience, the majority (1,271 or 79%) have received the fixed dose of 300 mg; 1,123 (69%) have been followed for at least 1 year, and 1,062 (66%) have been followed for over 2 years, satisfying, and indeed exceeding ICH safety database requirements [EMA Scientific discussion 2006].

Based on the clinical trial data, the risk of PML at that time was calculated as 1 per 1.000 after 17.9 month of treatment and appears to increase with the number of infusions. It should be noted however that at this point in time PML was not associated with monotherapy and could therefore be considered a potential risk rather than an identified risk.

Based on the CHMP review of data on quality, safety and efficacy, the CHMP considered by consensus that the risk-benefit balance of natalizumab indicated as single disease modifying therapy in highly active relapsing remitting multiple sclerosis for the following patient groups:

- Patients with high disease activity despite treatment with a beta-interferon (SPC: see 5.1);

or

- Patients with rapidly evolving severe relapsing remitting multiple sclerosis (SPC: see 5.1).

was favourable and therefore recommended the granting of the marketing authorisation.

A subgroup analysis in AFFIRM Major Efficacy Endpoints (Patients with ≥2 relapses and ≥1 Gd-enhancing lesion) in 148 natalizumab treated patients versus 61 Placebo treated patents confirmed this conclusion.

1.2.4.2 PML data at time of re-evaluation

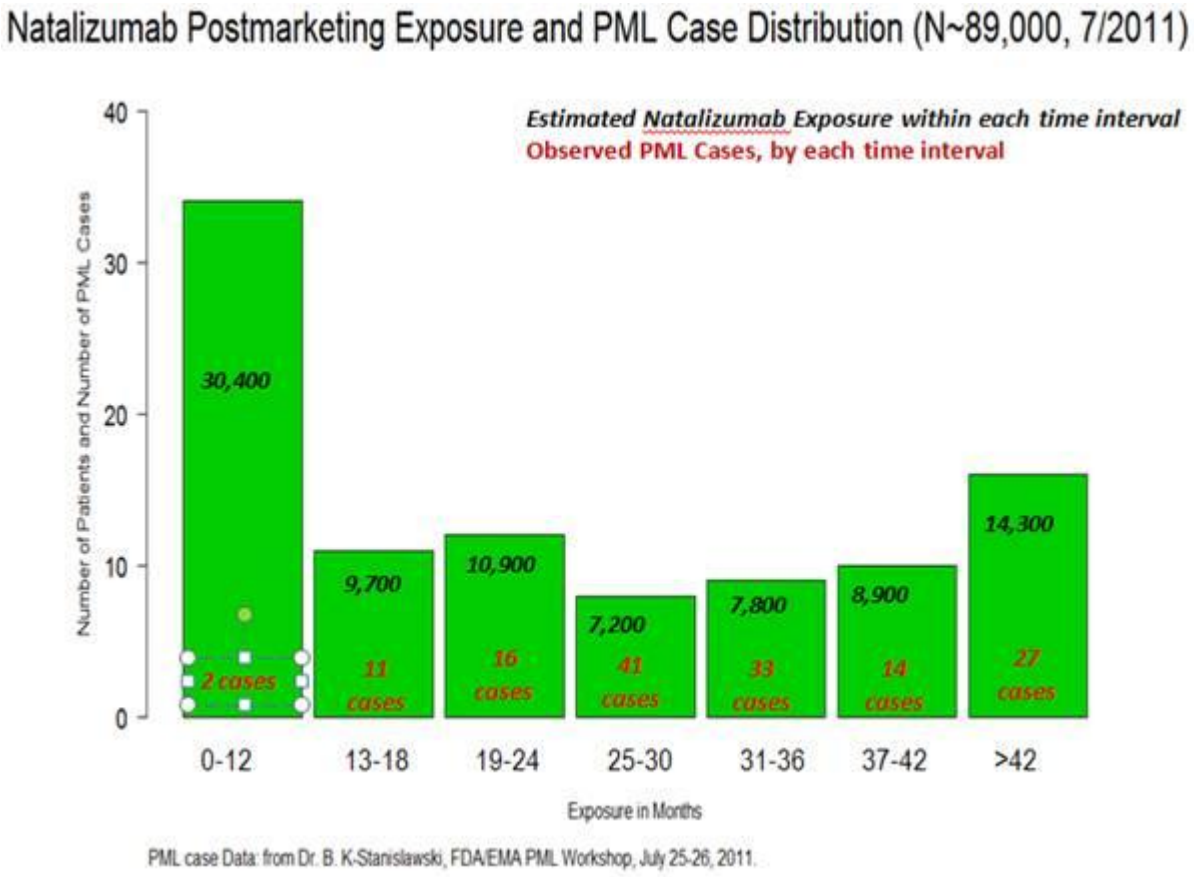
Twenty-three confirmed cases of PML had been reported worldwide in patients with Multiple sclerosis (MS) receiving natalizumab between July 2008 and October 2009, resulting in four deaths. With increasing post-marketing experience and duration of exposure to natalizumab, the continued reporting of MS patients diagnosed with PML raised concerns, especially because these data suggest, that the risk for developing PML increases significantly after two years of continuous exposure.

The CHMP reviewed natalizumab after it had received reports of side effects in patients receiving the medicine. As mentioned above, these included 23 confirmed cases of PML reported worldwide between July 2008 and October 2009, resulting in four deaths. 14 of these cases, including one death, were reported in the EU. By the end of the review procedure on 20 January 2010, the total number of confirmed PML cases had risen to 31 worldwide, of whom 23 had been receiving natalizumab for more than two years. This is equivalent to around one case of PML for every 1,000 patients treated with natalizumab for two years or more. It has been seen that PML case frequency increases over time. At present it is observed that the frequency of PML cases increases dramatically after 24 months (monthly infusions) of treatment (4).

As of May 4 2011 on the basis of post marketing reports, the estimated overall risk of PML has been reported as 1.51 per 1000 patients (95% CI 1.27-1.79) which is generally similar to rates seen in clinical trials (5).

More detailed observations were presented at a joint FDA / EMA workshop in July 2011. Figure 1.1 depicts in black, the number of patients exposed within each time interval, and in the red, the number of cases of PML observed with natalizumab, as of July 2011 (6).

Figure 1.1 PML post-marketing incidence



1.2.4.3 CHMP conclusion

Having considered the overall data provided by the MAH in writing and in the oral explanation, the CHMP concluded that the benefit still outweighs the risks for the patients treated with natalizumab. The CHMP also concluded that the Product Information for natalizumab should include safety information aiming at informing patients and physicians about the risk of PML so that the symptoms are detected as soon as possible and therefore recommended the amendments to the relevant sections of the Summaries of Product Characteristics and Package Leaflet.

The Agency's Committee for Medicinal Products for Human Use (CHMP) has concluded that the risk of developing PML increases after two years of use of natalizumab, although this risk remains low. However, the benefits of the medicine continue to outweigh its risks for patients with highly active relapsing-remitting multiple sclerosis, for whom there are few treatment options available.

Because it is important that PML is detected early, the Committee recommended that a number of measures be put in place to ensure that patients and doctors are fully aware of the risks of PML .

1.2.4.4 Patient involvement in decision making

Public domain information from the website of the Europeans Medicines Agency indicates that: "The CHMP requested the views of patients and consumers on medicinal products under evaluation. This occurred three times in 2009; for the medicinal products Onsenal, Prezista and Tysabri. In each case, selected patients' organisations, fulfilling the Agency's criteria for interaction, were asked to answer in writing a list of questions adopted by the Committee, and in some cases were subsequently invited to participate in the CHMP discussion on the issues. These initiatives gathered the experience and views of the organisations on certain aspects of the current use of medicines, and this information has been taken into account by the CHMP on the opinions subsequently adopted and has been

reflected in the CHMP assessment report" (7). It also indicates that natalizumab was an "example of patient participation in SAGs meeting, where very positive feedback on the patient contribution was collected (4).

Part 1.3 Summary of Wave 1 case study

A Wave 1 case study into the benefit-risk profile of natalizumab as a treatment for RRMS has been carried out as part of Work Package 5 (http://www.imi-protect.eu/results.shtml. The case study addressed two questions relating to the licensing of natalizumab:

1. Should natalizumab be given marketing approval at the time of first registration?
2. Should natalizumab be kept on the market given that increased episodes of PML were observed?

In order to address these questions, the Wave 1 case study employed a number of benefit-risk methods, as set out in Table 1. These were selected from an initial shortlist of methodologies identified by Work Package 5 for the Wave 1 case study applications (http://www.imi-protect.eu/results.shtml). Table 1 briefly discusses the reasons why the methods were chosen for this particular case study. The Wave 1 team was divided into two subteams; this allowed two sets of benefit-risk approaches to be tested in parallel.

Table 1 Benefit-risk approaches included for testing in natalizumab Wave 1 case study

Aspect	BRAT subteam	ProACT subteam	Justification
Descriptive guidelines	Benefit Risk Action Team (BRAT) framework	PrOACT-URL framework	PrOACT-URL qualitative framework as MCDA fits well with this framework. BRAT is also applied so the two frameworks can be compared.
Benefit-risk assessment frameworks	Benefit Risk Action Team (BRAT) framework	Multi-Criteria Decision Analysis (MCDA).	MCDA as this is an intricate method capable of integrating multiple benefit and risk-criteria. It is one of the more complex methods. BRAT is a framework supported by a set of guidelines and a tool that allows to structure data for decision making.
Metric indices	Number Needed to Treat (NNT) and Number Needed to Harm (NNH).	Benefit-Risk Ratio (BRR).	MCDA naturally leads to a benefit-risk ratio (or actually benefit-risk balance as a ratio is not generally used in MCDA). NNT and NNH is a popular method which is relatively simple to apply. It would be insightful to assess how it compares to more principled methods for assessing benefit-risk, in particular looking at the improvements made to this method in the recent past.
Estimation techniques	Mixed Treatment Comparison (MTC).	Mixed Treatment Comparison (MTC).	Use the same methods for both frameworks so data can be shared between them.
Utility survey	Direct elicitation	Direct elicitation	Use the same methods for both frameworks so weights can be shared between them. Direct elicitation is used

techniques due to time constraints.

Although the subteams employed different methodologies at certain stages of the assessment, there were elements common to both subteams, namely:

- identification of the decision context (the problem, objectives and alternatives)
- extraction of source data; and
- elicitation of preference information.

The Appendix describes how these tasks were dealt with in the Wave 1 case study. The extracted data on clinical outcomes and elicited preference weights are summarised in Table 2 below.

Using multi-criteria decision analysis, the methodology employed by the second subteam, for a particular treatment, the contribution of each outcome to the overall benefit-risk score is obtained by:

1. Determining where the observed outcome falls on the value scale. In Table 2, "Best value" means a score of 1 on the value scale and "Worst value" means a score of 0. Intermediate values are determined by interpolation, as value functions are assumed to be linear (with the exception of the value function for the route of administration, which is described in the Appendix). For example, the relapse rate of 1.04 for glatiramer acetate translates to a value of 0.48.

2. Multiplying the value score for each outcome by the weight of that outcome.

The overall benefit-risk score for each treatment is then obtained by summing the weighted value scores for all the outcomes.

Table 2 Endpoints, preference weights and clinical data used in the Wave 1 case study. The Appendix explains how the figures were derived.

Outcome	Expressed as	Best value	Worst value	Weight	Treatment	Quantity
Relapse	Average number of relapses per patient per 2 years	0	2	7.5%	Placebo Natalizumab Beta-interferon Glatiramer acetate	1.46 0.47 1.19 1.04
Disability progression	Probability of avoiding EDSS progression for 2 years	0%	100%	5.4%	Placebo Natalizumab Beta-interferon Glatiramer acetate	23.0% 11.3% 14.0% 18.2%
Reactivation of serious herpes viral infections	Proportion of patients experiencing this outcome in 2 years	0%	100%	6.5%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 0.0% 0.0% 0.0%
PML	Proportion of patients experiencing this outcome in 2 years	0%	100%	53.8%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 0.2% 0.0% 0.0%
Congenital abnormalities	Proportion of patients experiencing this outcome in 2 years	0%	100%	5.4%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 0.0% 0.0% 0.0%
Transaminases elevation	Proportion of patients	0%	100%	10.8%	Placebo Natalizumab	4.0% 5.0%

	experiencing this outcome in 2 years				Beta-interferon Glatiramer acetate	4.0% 4.0%
Seizures	Proportion of patients experiencing this outcome in 2 years	0%	100%	5.4%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 0.0% 3.0% 0.0%
Infusion reactions/injection reactions	Proportion of patients experiencing this outcome in 2 years	0%	100%	2.7%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 23.6% 17.6% 26.9%
Hypersensitivity Reactions	Proportion of patients experiencing this outcome in 2 years	0%	100%	1.1%	Placebo Natalizumab Beta-interferon Glatiramer acetate	0.0% 0.0% 0.0% 0.0%
Flu-like reactions	Proportion of patients experiencing this outcome in 2 years	0%	100%	1.1%	Placebo Natalizumab Beta-interferon Glatiramer acetate	39.9% 39.9% 60.8% 39.9%
Route of administration	Daily oral, daily SC, weekly IM, or monthly IV	Daily oral	Daily subcutaneous	0.5%	Placebo Natalizumab Beta-interferon Glatiramer acetate	Daily oral Monthly IV Weekly IM Daily SC

Both subteams concluded that, on the basis of the extracted data and elicited preference information, natalizumab should have been given marketing approval both at the time of first registration and at the CHMP's later re-evaluation.

Section 2 Aim and objectives

This document details an extension of the benefit-risk case study for natalizumab as part of the IMI PROTECT Work Package 5. Natalizumab (Tysabri®) was approved in 2004 by the FDA for the treatment of relapsing remitting multiple sclerosis (RRMS). In 2005 the drug was suspended because of an associated incidence of progressive multifocal leukoencephalopathy (PML), a rare neurological disorder. In 2006 it was re-introduced due to patient demand, but with strict risk minimisation measures. In 2011 the inclusion of anti-JC virus antibody status as a PML risk factor was included in the label of natalizumab (5).

In the EU, natalizumab was granted a Marketing Authorisation on 27 June 2006 for use as single disease modifying therapy in highly active relapsing remitting multiple sclerosis (RRMS) in patients with high disease activity despite treatment with a beta-interferon or in patients with rapidly evolving severe RRMS.

In the first wave of case studies two questions were addressed:

1. Should natalizumab be given marketing approval at the time of first registration?
2. Should natalizumab be kept on the market given that increased episodes of PML were observed?

Part 2.1 Key issues to be addressed

This extension to the natalizumab case study is aimed at addressing three key issues:

- Incorporate probabilistic uncertainty in the clinical data measures.
- Re-visit the weight elicitation and consider alternative methods for doing this.
- Use the case study to test the visualisation methods coming from the visual review work stream of WP5.

The case study team was divided into three sub-teams, each addressing one of the above key issues, as shown in Table 3 below.

Table 3 Team structure

Project	Person	Areas of expertise
Weight elicitation	**Kimberley Hockley** Alain Micaleff Adam Elmachtoub Richard Nixon	Patient and public involvement Physician Operations Research Statistics, Decision analysis
Probabilistic uncertainty	**Ed Waddingham** Richard Nixon Nan Wang	Statistics Statistics, Decision analysis Statistics
Visualisation	**Shahrul Mt-Isa** Isabelle Stoeckert Silvia Kuhls Dave Gelb Andrew Thomson Rick Hermann	Statistics Physician / Regulatory Affairs Biostatistics – Integrated Analysis Statistics (SAS, Spotfire, Stata) Statistics (R) Physician

The weight elicitation work involves directly eliciting preferences from a sample of people who have been affected by RRMS. Due to unforeseen delays in recruiting this sample, the results of this work are not yet available. The full details of this part of the project are therefore not included in this report but will be made available in a separate report.

Section 3 Visualisation methods

Part 3.1 Introduction

A report has been written as part of IMI PROTECT WP5 on visual methods for use in benefit-risk analysis based on commonly used visuals encountered in the PROTECT Methodology Review (8). The first wave natalizumab case study had adopted some of these common visual displays as well as experimented with other suitable visual with the aim to improve the communication of benefits and risks. This report systematically explores the methods for creating visualisation with the focus to make recommendations for their use in practice.

Part 3.2 Methods

The aim of this part of the case study is to review and critic visual methods that were used in the Wave 1 natalizumab case study, but it is not our intention to invent new visual displays. We worked closely with the main PROTECT visual review team to identify alternative visualisations from the literature and other sources which may be applied in our work.

There are three phases involved in the work of visualisation method sub-team:

(1) Phase I: to review and appraise visual methods that were used in the first wave of the case study
(2) Phase II: to investigate creation of animation and interactive features of reviewed graphics
(3) Phase III: to guide and critic visual methods used in the stochastic uncertainty sub-teams, considering suggestions from PROTECT Visual Review team

We summarise the results from Phase I and II in Sections 3.3.1 and 3.3.2 respectively, and present more detailed outputs from Phase I as trading cards ("baseball cards") in Appendix Part 7.4, and further document the results from Phase II in Appendix Part 7.5 respectively. We present the outputs from Phase III within the respective sections (Section 4 and Section 5).

3.2.1 Phase I: Graphics review and appraisal

Phase I of the sub-project is the most straightforward although the most crucial and was very time consuming. In this phase, the natalizumab Visual Methodology team split into three smaller groups (Bayer Group, Takeda Group, and Imperial + others) to review and appraise the graphics that were produced in the Wave 1 case study. The Bayer Group had also invited three statisticians and one regulator colleagues within the company but external to PROTECT to assist in the review and appraisal through a face-to-face meeting. The other two groups consist of one member each but all appraisals were brought back for the whole team's discussions and comments.

The rationale to formally appraise our own work is to learn and understand our initial judgment of presenting data from past experience. To facilitate the appraisal process, we articulate the underlying aspects of the graphics using the criteria given in Table 4. At the end of Phase I, we select a sample of graphics to be used in the case study, and would benefit from some improvement, particularly to be translated into interactive graphics.

Table 4 Criteria for assessing visual displays

	Short name	Description
(a)	Name/rubric	What is the visual/graphic type/name? These are already grouped by graphic type. *But is there a special name for the visual?*
(b)	Created in	Which technology is needed to produce the visual? *This is the software that was used to create the visual.*
(c)	Message	What does the visual communicate? *This is the intended main message(s) from the visual. Can the benefit-risk quantity be determined?*
(d)	Intended audience	Who is the intended audience? *The main user(s) the visual is aimed at.*
(e)	Knowledge required	What do I need to know to understand the visual? *This is the background*

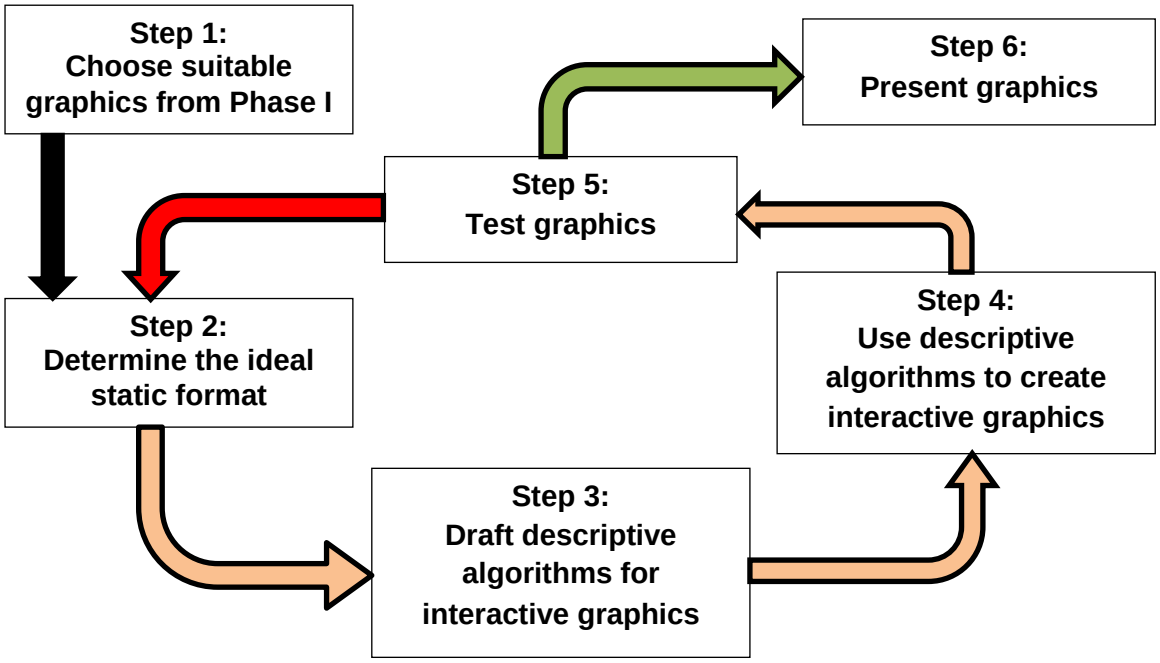
		knowledge of the intended audience specifically, and typical knowledge of other audience in order to understand the main message(s) clearly.
(f)	Unintentional message	Does the visual unintentionally convey/communicate a confusing message or a message that could be misunderstood? *This is to address any misleading or confusing information on the visual that may be due to the visual display design, or lack of user's knowledge. Appraisal should address:* - *does the visual reflect the original data?* - *is there any misleading assurance of the benefit-risk balance?* - *how does the visual reflect the amount of certainty/uncertainty of the benefit-risk balance?*
(g)	Message not communicated	What does the visual not convey/communicate? *This should be addressed within context of the above.* - *is benefit-risk balance communicated clearly, only one dimension, or both dimensions but not integrated?* - *is there implicit message that could be made clearer in design or become clearer with improved knowledge?* - *is there any known weakness of the visual?*
(h)	Proposed improvement	*Any proposed improvement that can be made to the visual to make it more useful?*

3.2.2 Phase II: Graphics animation and interactivity

Following the review and appraisal in Phase I, we focus our effort on improving and creating interactive versions of a selected set of graphics following the guidelines and recommendations from the PROTECT Visual Review team. Figure 3.1 demonstrates the six steps we have taken to complete Phase II. The aim of Phase II is to demonstrate the feasibility and as proof of concept to encourage more active stakeholders participation in benefit-risk assessment of medicines via graphic interactivity. Phase II delivers a set of easy-to-use and adaptable templates to create the interactive graphics as demonstrated.

Step 1 follows directly from Phase I where we have selected a number of graphics to be improved further. The judgment as to which graphics is based on the team's subjective view on the graphics usefulness and attractiveness for use in this case study. We make brief justifications of the reasons for taking certain graphics into Phase II in Appendix Part 7.4 at the end of the appraisal tables.

Figure 3.1 Visual methodology Phase II workflow



We determine the ideal static format of a graphic in Step 2. The ideal static format of the graphic is the first image being displayed to the intended audience when viewing the interactive version of the graphics, or one that appears in printed form. It is therefore important that the ideal static format characterises and delivers the message efficiently. The ideal static format is constructed in reference to the principles in the GSK Graphics Principles – published online on as part of the CTSpedia Safety Graphics Wiki site (https://www.ctspedia.org/do/view/CTSpedia/BestPractices). The graphic as displayed in the Wave 1 case study may already be the ideal static format, so may not require much improvement. The GSK Graphics Principles are reworked into five simple criteria to be addressed in Step 2:

(1) Communication – what is the primary communication purpose of the graphic? The primary communication purpose should be intuitive and provide clear insight into the message.
(2) Annotation – how should the texts on the graphic appear? The texts and other annotations such as legends should be legible and in suitable fonts taking the format of presentation into account (printed on paper, display on computer screen, projection in conferences etc.). The annotations should help interpret the graphic and position appropriately so not to distract the main message.
(3) Axes – how should the axes appear (including labels and axis titles)? Axes should support the data they are presenting and aid the interpretation of the graphic (for more than one graphs, axes scales must be the same for meaningful comparison).
(4) Styles – what are the symbols or line patterns to use on the graphics? Symbols and lines should be distinct and readable taking into considerations the format of presentations. They should also be familiar and intuitive where possible and consistent across similar graphics.
(5) Colours – which colours are to be used to encode different entities on the graphics? The colours should be contrasting and clearly visible, as well as compatible with the background. The design should also take into account colour combinations that are suitable for people with colour deficient vision. Variation of colour can imply grouping and similarity; therefore should only be introduced as necessary. Color Brewer is an excellent reference for choice of colours (http://colorbrewer2.org).

Step 3 of Phase II takes the static graphics into a new dimension by adding elements of interactivity into the graphics. Our approach to this task is to create a descriptive algorithm for each graphic in preparation of constructing the interactive version. The descriptive algorithm plays two roles: (1) to provide a set of instructions, along with the suggestions from Step 2 above, to the graphic designer on how to create the interactive version of the graphics, and (2) to allude users to the interactive features of the graphic. The descriptive algorithms address the following five items:

(i) Specify the main intended audience. For logistic purposes, we assume minimal technical knowledge requirements.
(ii) What is the first image to show? This would be the ideal static format from Step 2.
(iii) Which variables/elements on the graphic would benefit from filtering? We would like to find out whether some elements on the image could be temporarily hidden or highlighted, and whether the filtering should be done at the data level to enable subgroups of, say, patients with certain characteristics be displayed.

Which are the important elements that would benefit from animation or interactive capabilities? The specifications included are which elements to animate, the order of animation (can also be simultaneous), how should the elements move (c.f. Wickens' principle 7: Moving part – *The moving elements of a dynamic display should move in spatial and direction that are compatible with how the users think they actually move in physical system* (9)).

(iv) Which elements on the graphic could be interrogated further or be given drill-down facility? An example of drill-down is by clicking on a total score bar (on a bar graph) to open up its components. This could be done by changing views (show a different graph on a different page) or by expanding/collapsing the current view.

In Step 4, we use the suggestions and instructions from Step 2 and Step 3 to create interactive versions of the graphics. Our primary outputs were created using the Tableau Public software since we have access to the software and managed to quickly grasp its concepts and functionality (http://www.tableausoftware.com/public/community).

Since the creation of graphics that are suitable for use is rarely done as a one-step process, Figure 3.1 indicates a cyclic workflow, in which the interactive graphics created are tested in Step 5 by the sub-team and improved further until no significant improvements could be made. Step 5 is conducted through a telephone meeting among case study team members following emails communication on the matter. The final Step 6 is to present the graphics that the sub-team are satisfied with and agreed upon numerous discussions. The final set of graphics is presented Appendix Part 7.5. The intermediate graphics are not shown or described further in this report as they are regarded as "not fit for purpose".

We have not documented our approach to the technical aspects for creating the interactive graphic templates using Tableau Public, but the worksheets are freely downloadable from the links to the graphics, and much easier to be learned by examples.

3.2.3 Phase III: Working with other case study sub-teams

In Phase III, we use the templates created for interactive graphing to create appropriate graphics using the data available related to probabilistic uncertainty analyses. We also identify the respective sub-team's visualisation needs resulted in further templates to be generated to support the data (see Section 3.3.3).

Part 3.3 Results

The results from the visualisation methodology work are structured according to the three phases described above. In this section, we summarise the key points and general overview of the results from the graphics review and appraisal (Section 3.3.1), the graphics animation and interactivity (Section 3.3.2) and the visual work within the other sub-teams (Section 3.3.3). The appendices Part 7.4 and Part 7.5 serve as the presentations of the detailed documentation for Phase I and Phase II respectively. The outcome of the visualisation methodology in Phase III are presented and discussed within the sections on Probabilistic uncertainty (Section 4); whilst Section 3.3.3 only provides very brief account on the visualisation aspects that took place within the other sub-teams.

3.3.1 Graphics review and appraisal

The review and appraisal of graphics from Wave 1 of natalizumab case study reveal three areas of importance that are often not given enough considerations and would benefit from greater attention. Full details of the review are available in Appendix Part 7.4 and it contains appraisals on the following graphics:

a) Flow charts (Appendices 7.4.1 and 7.4.2)
b) Tables (Appendices 7.4.3 to 7.4.5)
c) Line graphs (Appendices 7.4.6 to 7.4.8)
d) Value trees (Appendices 7.4.9 and 7.4.10)
e) Dot/Forest plot (Appendix 7.4.11)
f) Bar charts (Appendices 7.4.12 to 7.4.17)
g) Thermometer scale (Appendix 7.4.18)

One area is in the visual representation of graphics itself. The key messages to take home are:

- Categorical outcomes are to be presented using visual display method for displaying discrete variables e.g. bar graph, dot plot or symbols. Categorical outcomes should not be displayed as line graph as the categorical outcomes would be perceived as being continuous and that the connecting lines are meaningful when in fact they are not.

- Preference values and weights on benefit-risk criteria should be presented alongside the probability or frequency data on the criteria since preference values and weights are prone to misinterpretation when displayed alone.

- The labelling of criteria should be intuitive and clearly describe whether a criterion is a benefit or a risk. The labelling should also give insight to the direction of the measure. For example, the use of "reduction in relapse rate" would better inform the decision-maker instead of "relapse" to convey that the criterion is a benefit. However, care should be taken in the case of terms like "relapse" and "disability progression" since the actual data represents the number of relapse and number of those who had disability progression. Therefore the introduction of the term "reduction" which already indicates a negative effect to describe a benefit may be confusing and inaccurate when presented with the data (the effect of "double-negatives")

- Rare events should be given more careful considerations given the circumstances when they are presented to avoid underestimation, or in some cases overestimation. Very small risks could appear negligible but may be regarded by the regulators as very important and a key decision driver.

- The presentation of the graphics should be self-contained and explanatory without long supporting texts. This could be achieved by using familiar graphics and benefit-risk metrics, as well as by avoiding packing too much information into one graph when it is better done as a series of graphs.

- The use of absolute number can support judgments and should complement relative data.

The second area that requires attention relates to the appearance of the graphics. The key messages to take home are:

- Colouring of criteria should reflect groupings and should be meaningful – benefits, risks, etc. Avoid introducing meaningless variations of colours.

- Benefit criteria on a graph should not be presented using the same colour as risk criteria to avoid confusion since entities with the same colours imply grouping.

- The red/green scheme to encode risks/benefits respectively appears to be preferable to other colour combinations. Because of the familiarity, the red and green should be reserved for risks and benefits to avoid confusion. However, the red/green scheme may not be suitable for people with colour deficient vision, and thus alternative colour scheme should be used (see http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer_intro.html for alternative colour palettes.

- More attention should be given to the axes that carry the crucial information. This includes labelling and naming the axes, choosing the best aspect ratios to represent the relationship between horizontal and vertical axes, and choosing and displaying meaningful range and scales to reflect the data.

- When presenting quantitative measures of the benefits and risks, the preference appears to be to sort by magnitude to help judgment when there is no clear ordering.

Finally, the third area we found is that graphic labelling has not been emphasised enough. The key messages to take home are:

- All labels should be legible to the users. The characteristics of the intended audience and the format of presentation, whether the graphics are presented on a poster, printed on paper, projected to a screen, viewed on a small or large computer screen etc. The colour of labels should be in black unless the colouring is meaningful to the interpretation of the graphics e.g. to draw attention to an entity. The Juice's Simple Font Framework could make deciding on fonts easier, and is available on http://www.juiceanalytics.com/writing/simple-font-framework/.

- The labels should be clear and intuitive with carefully chosen intervals, range, scales and orientation. As a general rule of thumb, always use horizontal labels for better readability.

- Graphic and axes titles should be meaningful and reflect the information they carry. Axes using log-scales should clearly state the unit. The position of the graphic title should be at the top of the graphic since it is the most common practice defaulted in various software packages. Vertical axes titles should appear in centre position (across its range) of the corresponding axis.

- Any text descriptions and labelling must use meaningful and non-confusing texts. Where appropriate, avoid labelling a measure with "higher" and "lower", and use "favour", "better", "worse" etc. as labels to allow easier and more helpful coding of information that would directly assist the decision.

- Unless restricted by regulations or for blinded assessment, the drug options must be named explicitly e.g. "natalizumab vs. placebo", not "natalizumab vs. comparator".

- Any instructions, notes and annotations on a graph must be clear and their roles should be understood. They must not be confused as data.

At the end of the review and appraisal process, the team has chosen six graphics to be taken forward for improvement. The selected graphics are value tree (Figure 3.2), discrete data representation (Figure 3.3), waterfall plot (Figure 3.4), aligned bar chart (Figure 3.5), tornado plot (Figure 3.6), and forest plot (Figure 3.7)[1]. Further details of the review and appraisal including a brief explanation of the choices are given in Appendix Part 7.4.

Figure 3.2 Value tree

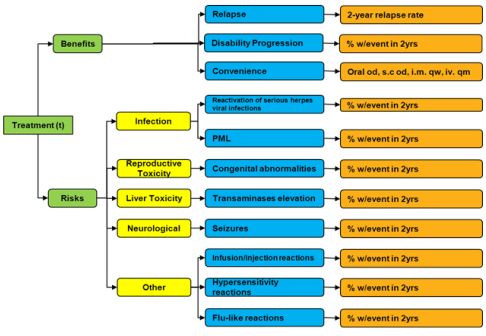
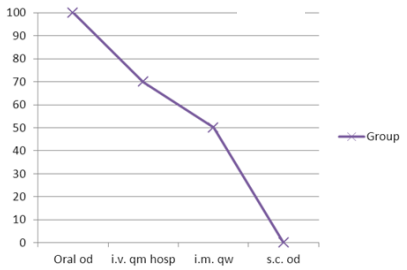


Figure 3.3 Value function for convenience



[1] If viewed in Microsoft Word, double-click Figure 3.2 to Figure 3.7 to view larger images of the graphs.

Figure 3.4 Incremental benefit-risk of natalizumab compared to placebo, Beta-interferon and Glatiramer acetate represented as waterfall plot

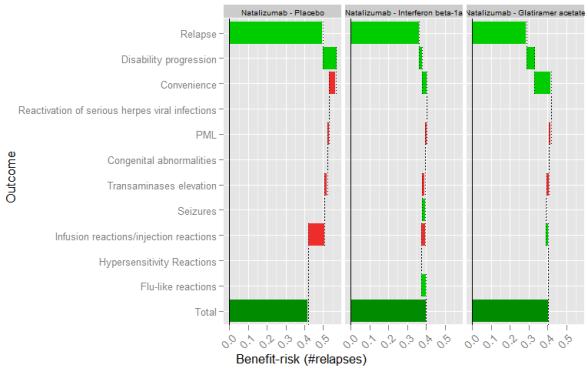


Figure 3.5 Benefit-risk of natalizumab compared to placebo

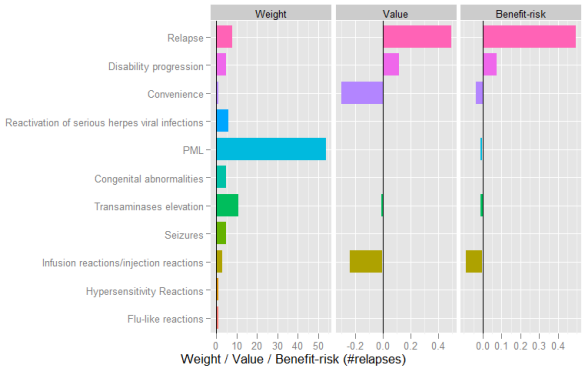


Figure 3.6 Tornado plot to assess how changes in the natalizumab outcome measure affect the incremental benefit-risk score

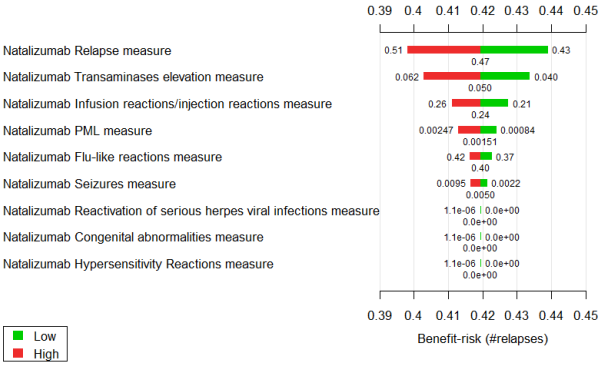
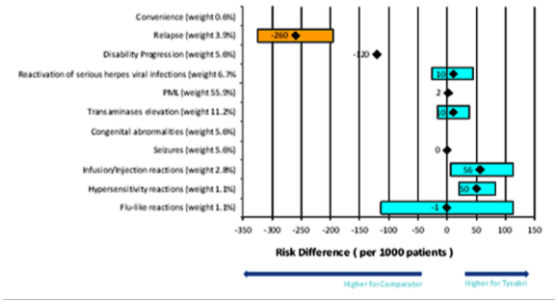


Figure 3.7 natalizumab versus placebo (comparator) at the time of CHMP re-evaluation

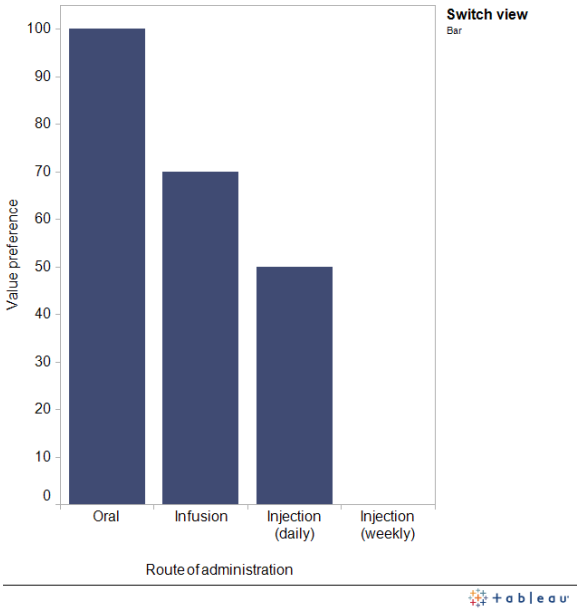


3.3.2 Graphics animation and interactivity

We discovered through the graphics review and appraisal that the major limiting factors of static graphical representations are the user-preference of visual display, insufficient information being displayed (not a self-contained graphic) and in many cases, the use of abstract or unfamiliar concepts on the visual display. These factors contribute to the ineffectiveness in the communication of important benefit-risk messages on these visual displays. These issues are addressed as much as possible in Phase II when generating the interactive graphics.

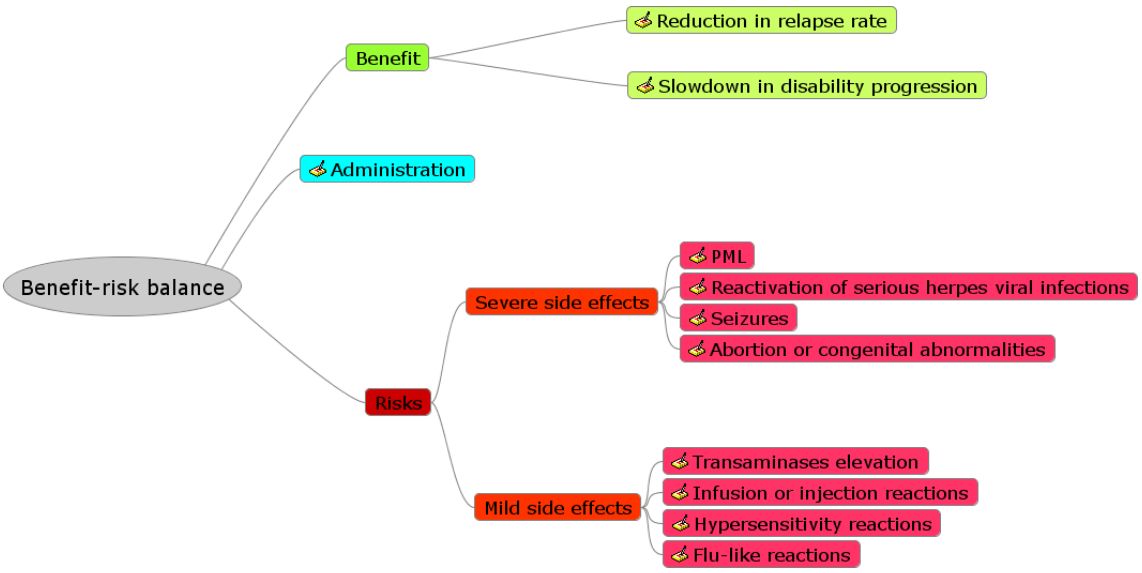
The six graphics were created in Tableau Public software (http://www.tableausoftware.com/public) and are briefly summarised here in Figure 3.8 to Figure 3.15. When designing the visual displays, particular attention was given to the type of data being represented. The original plots shown in Section 3.3.1 respect the data types with the exception of Figure 3.3 where discrete data are presented as if they were continuous.

Figure 3.8 Bar graph representing discrete data



Discrete data are presented in two ways in Phase II: as a simple bar graph (Figure 3.8) and as a dot chart with short line bar to replace the dot symbol (Figure 7.3 in the Appendix). Both visual displays clearly show that the data are discrete with the latter having higher data-to-ink ratio. In the latter representation, it is also possible to add information on uncertainty such a confidence interval line without over-cluttering. The interactive version of the discrete data representation graphs is available at http://public.tableausoftware.com/views/T_Discrete/Discrete where users can choose their preferred representation between the simple bar and the short line bar graphs.

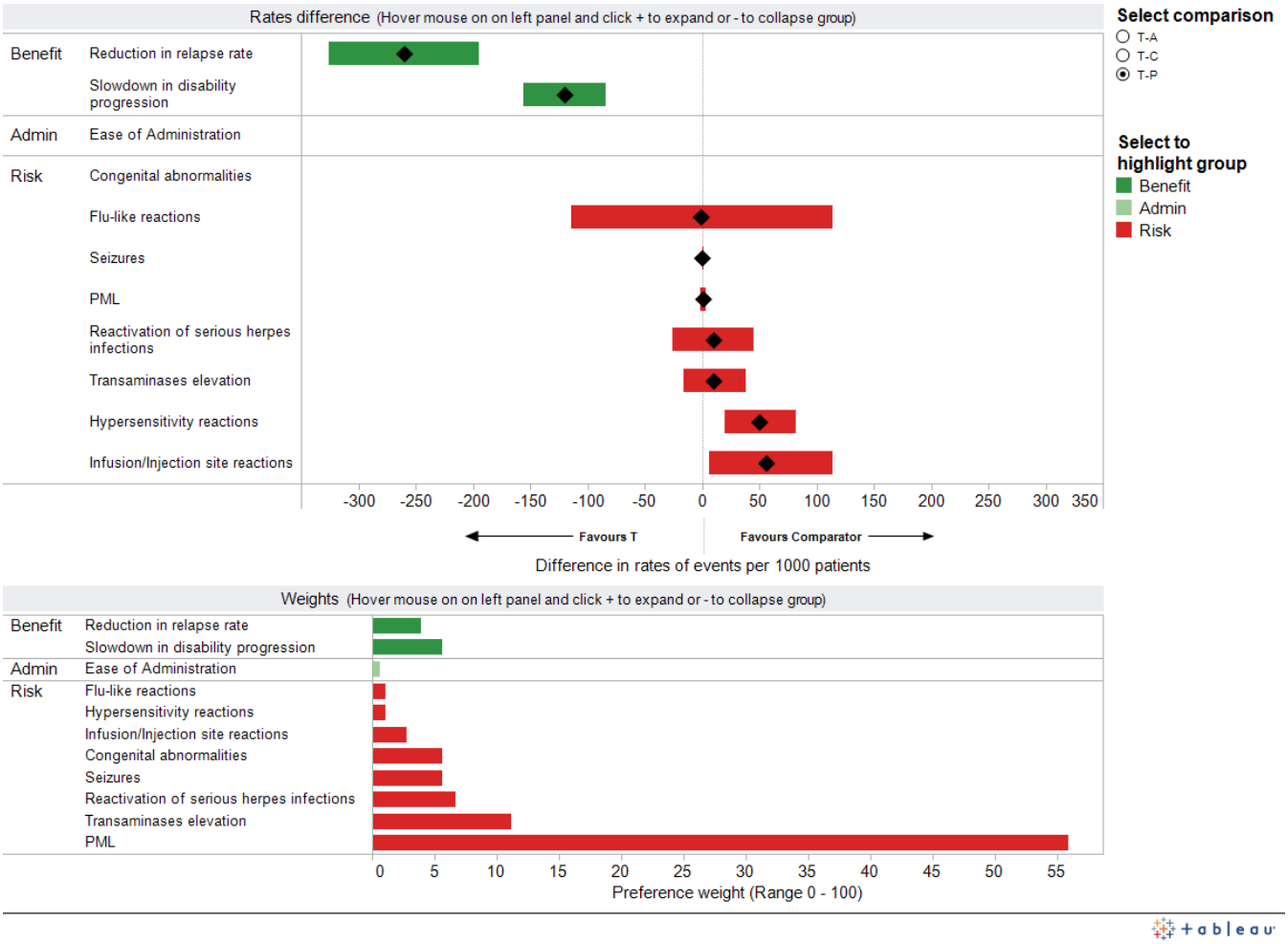
Figure 3.9 Tree diagram showing the criteria used in the model and their hierarchy



The value tree is represented as a tree diagram (Figure 3.9). It was agreed that tree diagram is the best form of representation for a value tree to show the hierarchical nature of the benefit-risk criteria. Since the value tree is adapted for use with patients, colour-coding and generous annotations were added to the diagram to induce interest and to aid understanding. We recognise that introducing colour into the tree diagram violates the data-to-ink ratio but it also adds to redundancy gain (9) where a user can quickly see the criteria in benefit and risk group. However, the colour may need to be made lighter because they may appear darker when projected onto a screen.

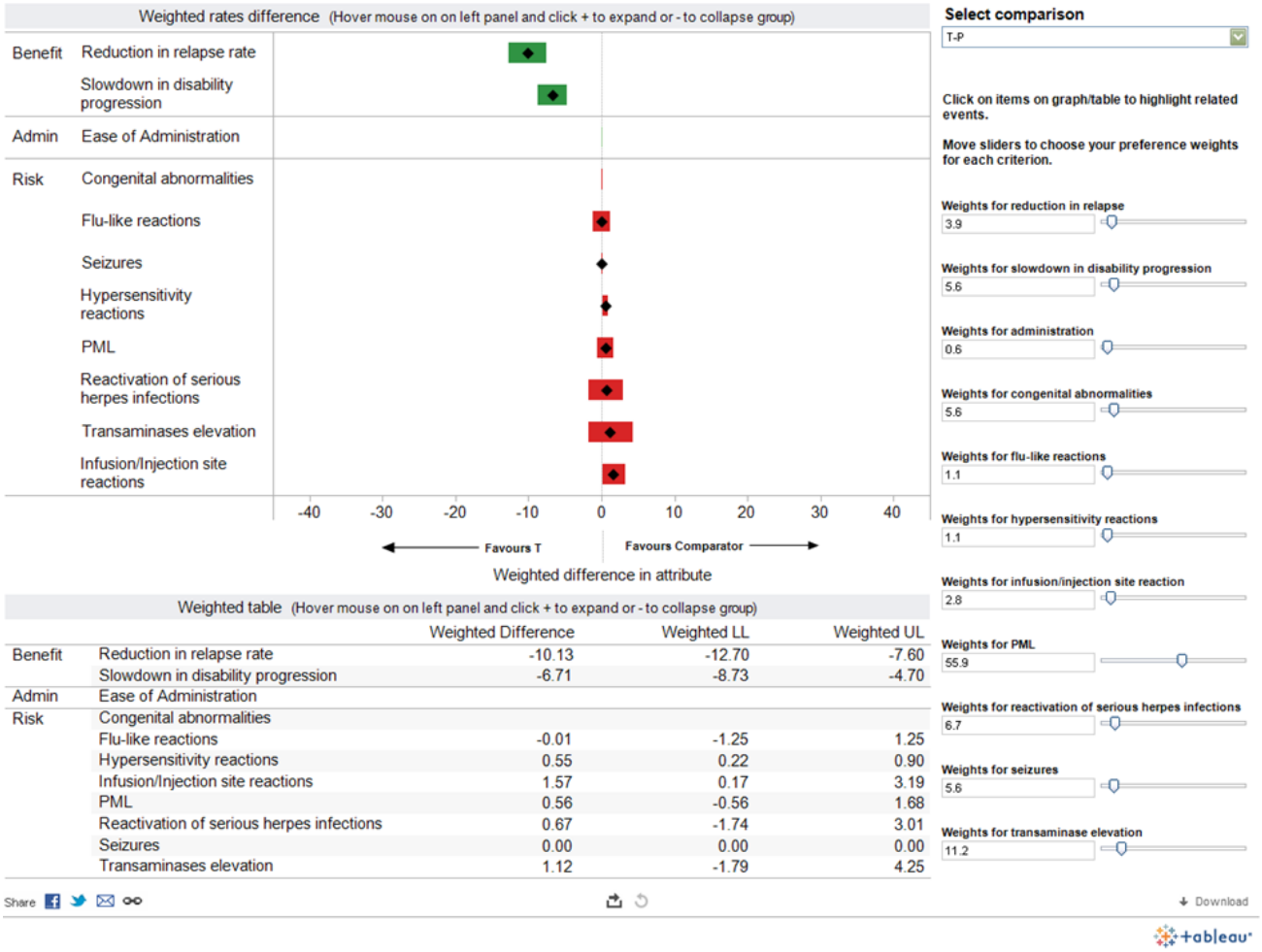
The annotations appear in the interactive version of the tree diagram that was created, and are listed in Appendix 7.5.2.

Figure 3.10 Forest plot with bar graph of preference weights



A forest plot is essentially a combination of a dot plot to represent the mean values and a bar graph to represent the confidence intervals (uncertainty). A forest plot of the rates difference (Figure 3.10) and another of the weighted difference (Figure 3.11) are created to complement each other. The forest plots were created as dashboards, where more than one graphic is on display simultaneously. The first dashboard (Figure 3.10) contains the rates difference as forest plot and a simple horizontal bar graph to represent the elicited patient preference weights.

Figure 3.11 Forest plot with weighted outcome measures



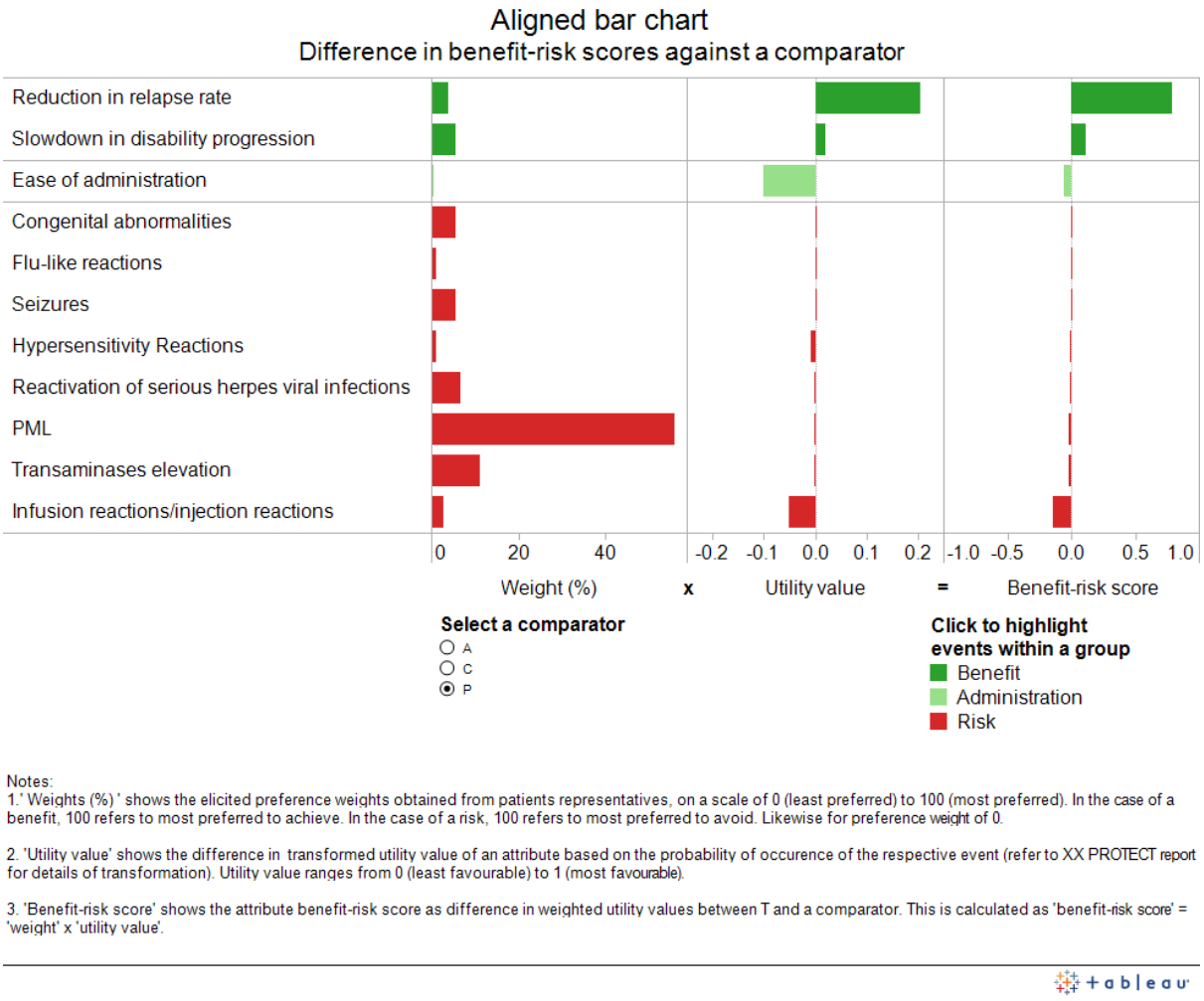
Weighted table (Hover mouse on on left panel and click + to expand or - to collapse group)

		Weighted Difference	Weighted LL	Weighted UL
Benefit	Reduction in relapse rate	-10.13	-12.70	-7.60
	Slowdown in disability progression	-6.71	-8.73	-4.70
Admin	Ease of Administration			
Risk	Congenital abnormalities			
	Flu-like reactions	-0.01	-1.25	1.25
	Hypersensitivity reactions	0.55	0.22	0.90
	Infusion/Injection site reactions	1.57	0.17	3.19
	PML	0.56	-0.56	1.68
	Reactivation of serious herpes infections	0.67	-1.74	3.01
	Seizures	0.00	0.00	0.00
	Transaminases elevation	1.12	-1.79	4.25

The combination of these two graphics is shown on the second dashboard (Figure 3.11) showing the weighted differences and a summary table of the results. In both dashboards, we added a filter function where users can select different comparators to natalizumab from a choice of placebo, beta-interferon or glatiramer acetate. Controls for user preference weights were also added to allow user to input their own preference weights to obtain more relevant results.

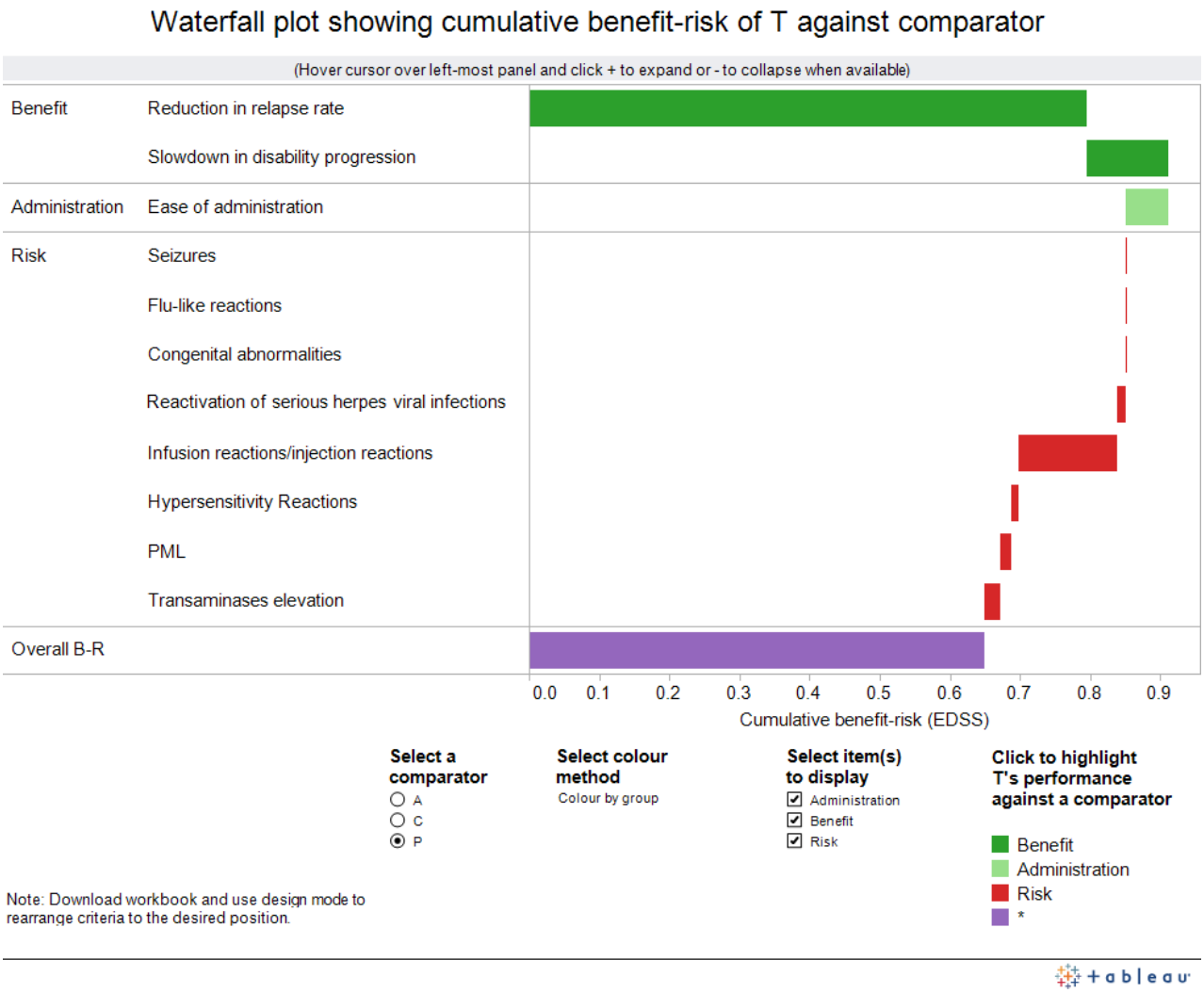
It is common practice to indicate the direction of drug preference on forest plots but it is quite challenging to be implemented in Tableau Public. We have added this feature in as blank graph underneath the forest plot which had to be carefully aligned resulted in having to fix the scale. This has the downside of diminishing visibility and discriminability when the magnitudes of the criteria are relatively small compared to the scale range. This is considered the limitation of the current software but could be easily rectified in different software. The interactive version of the first dashboard is available online at http://public.tableausoftware.com/views/T_Forest/ForestAndWgt and that for the second dashboard is http://public.tableausoftware.com/views/T_Forest/WgtNCB.

Figure 3.12 Aligned bar chart showing the weight, utility values and benefit-risk score by criterion



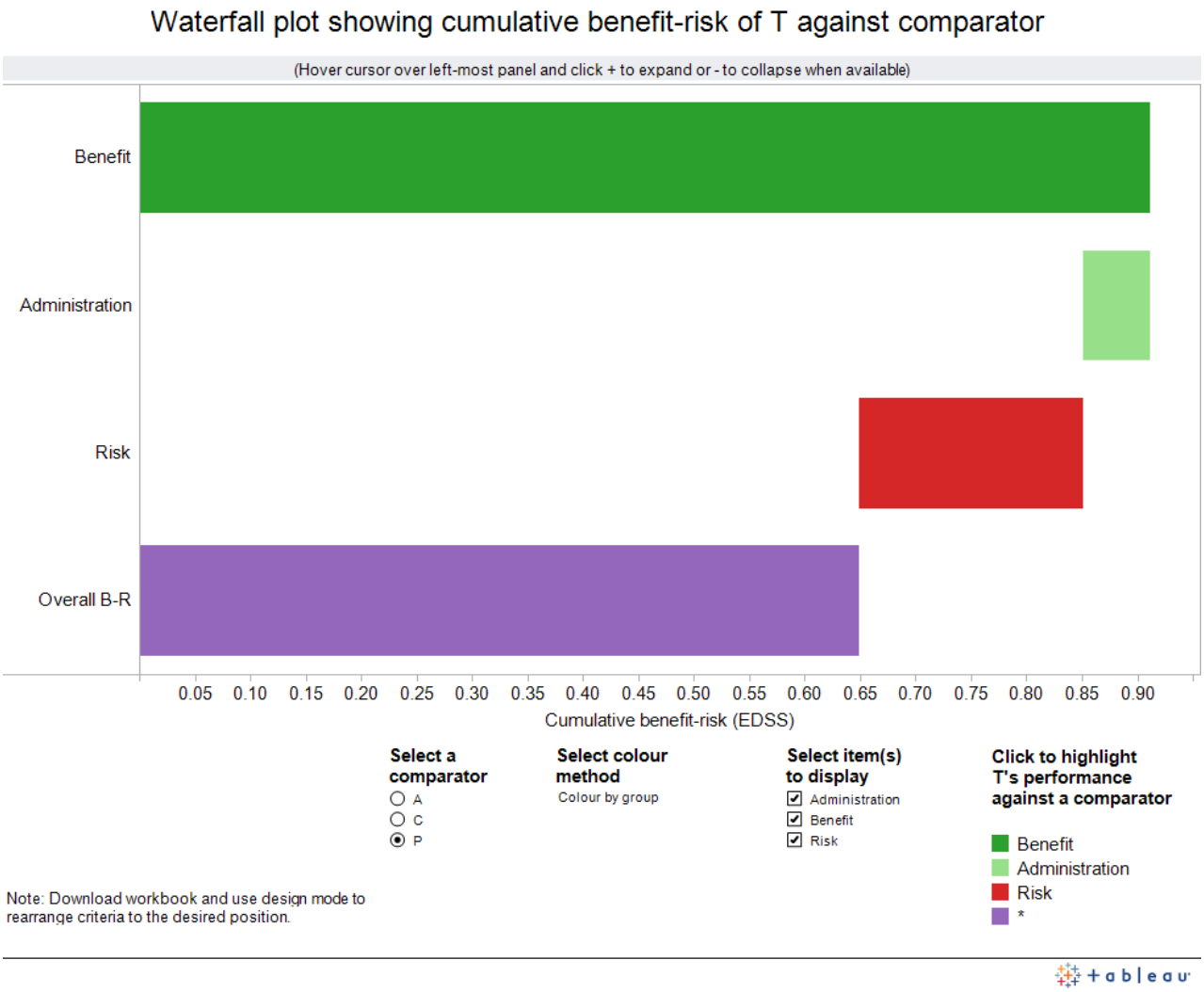
We continued using similar recipe as forest plot above in terms of colour-coding and filtering when creating the aligned bar chart (Figure 3.12). “Ease of administration” is regarded as a “benefit” and is coded with lighter shade of green. Grouping was done by high-level criteria group (benefit, administration, risk) to indicate hierarchy and to allow for within-group ordering of criteria. We added brief notes at the bottom the aligned bar graph to explain the three columns of the graph. The online interactive version packs more features and is available at http://public.tableausoftware.com/views/T_AlignedBar/AlignedBar.

Figure 3.13 Waterfall plot colour-coded by high-level criteria (group)



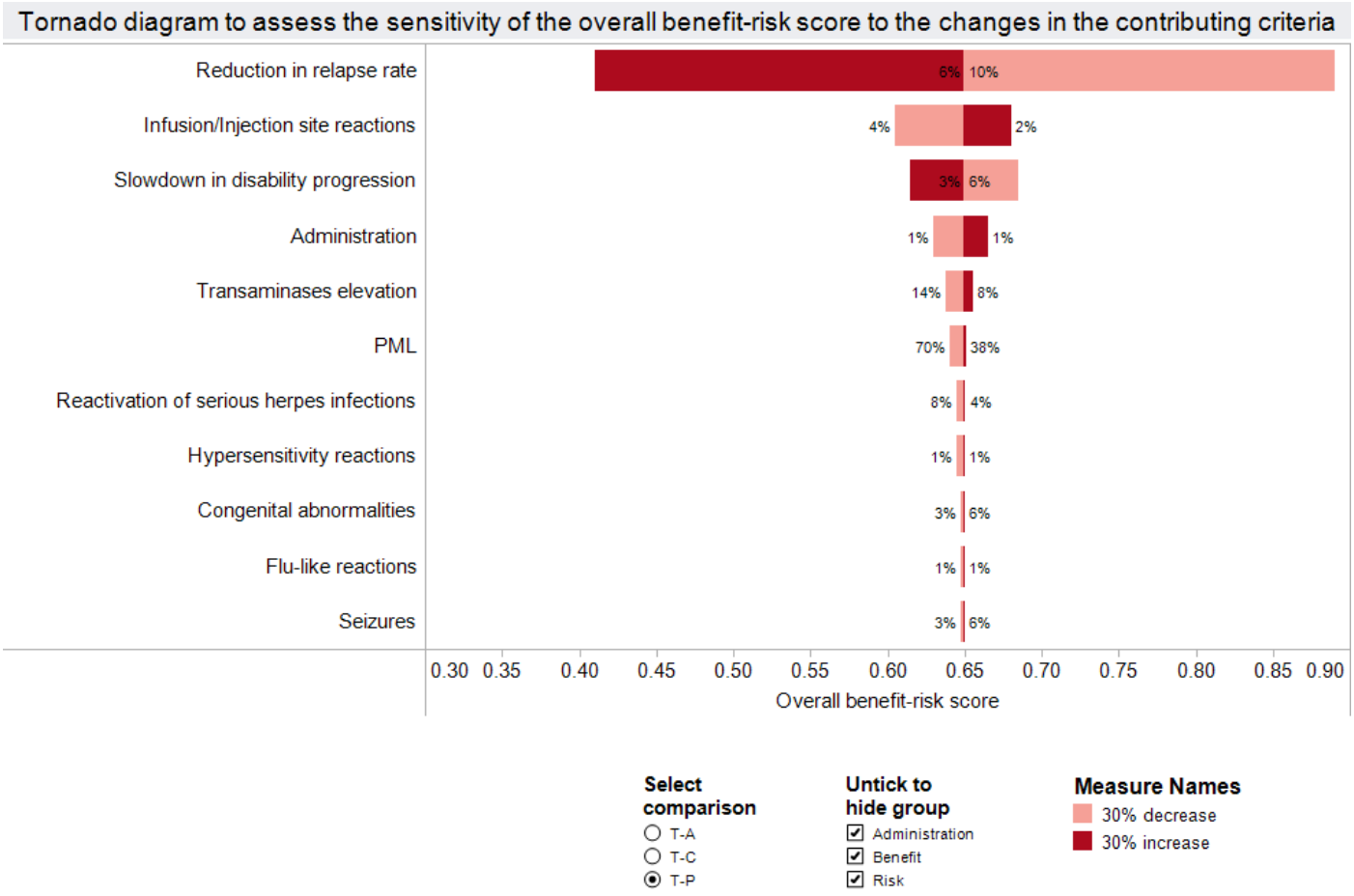
The waterfall plot shows the magnitudes and directions of each criteria and their impact on the cumulative benefit-risk score. Obviously, the overall appearance of the waterfall plot would be different if the criteria were ordered differently. This affects the running sum of the benefit-risk score but the overall benefit-risk score remains the same. We have colour-coded the bars in two ways to cater for users' preference to which information they would like to see. The first colour scheme is to colour bars by group of criteria whether they are benefit, administration or risk criteria; and the second is to colour by direction of effects whether a criterion increases or decreases the overall benefit-risk score. A comparator can be chosen from either placebo, beta-interferon or glatiramer acetate. Users may also choose to hide or display a group of criteria using the tick boxes available on the dashboard.

Figure 3.14 Drilled up waterfall plot of Figure 3.13



The individual criterion can be collapsed (drilled-up) into high-level groups to show the total contribution by group (Figure 3.14). The interactive version of the waterfall plot is available online at http://public.tableausoftware.com/views/T_Waterfall/WaterfallRisk.

Figure 3.15 Tornado diagram to assess the sensitivity of benefit-risk score to changes in the underlying criteria comparing natalizumab to a comparator



Notes:
1. The 30% value to analyse the sensitivity for change in score is was determined arbitrarily.
2. Preference weights incorporated into the analyses were those of patient representatives elicited through a decision conference.
3. Evidence data on the benefits and risks events were obtained from published clinical trials.

The tornado diagram from the first wave of the case study was replicated with slight amendments. We applied similar filtering in the interactive version to allow users to select comparator to natalizumab from beta-interferon, glatiramer acetate or placebo. There is also a filter to hide a group of criteria, as in the waterfall plot, should the users chose to. Colour-coding identifies the direction of change that is whether the changes in overall benefit-risk score are attributable to an increase or decrease in a criterion's base value. The tornado-like effect is achieved by ordering the criteria by the magnitude of change in the overall benefit-risk score corresponding to the maximum and minimum criterion values. Unavailable data are labelled as "indeterminate" at the base overall benefit-risk score. The interactive version of the tornado diagram is available at http://public.tableausoftware.com/views/T_Tornado/T_Tornado.

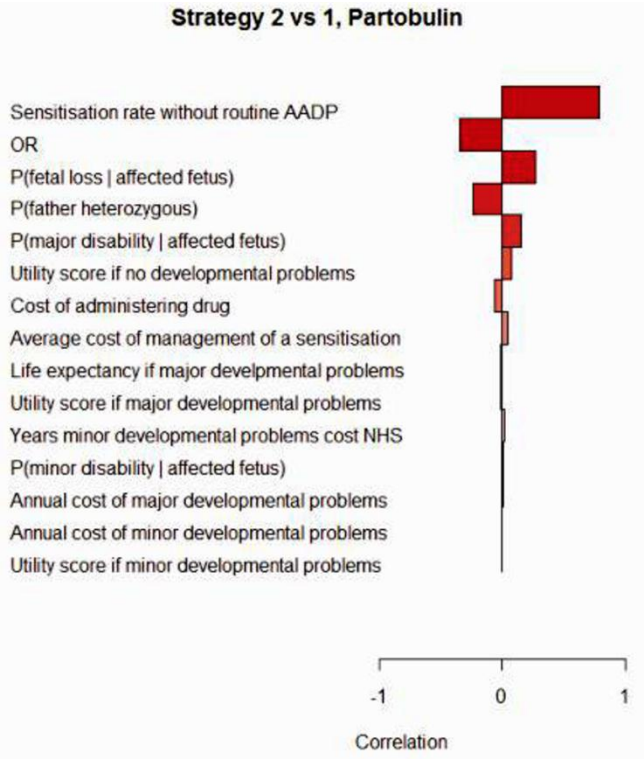
3.3.3 Working with other case study sub-teams

So far the visualisation methodology sub-team has only revisited the graphics produced in the Wave 1 of PROTECT natalizumab case study, as we work in parallel to other sub-teams. As previously mentioned, we aimed at producing

interactive graphic templates for the six graphics that were investigated in Phase II which are easily adaptable for use with new data (see data structure in Appendix Part 7.5).

In the probabilistic uncertainty sub-team, we identified that visualisation methodology to show the uncertainties in the form of statistical distributions is required and is of special interest. Statistical uncertainty in parameters are typically shown as distribution plots, box plots and also as forest plots as used in Wave 1 case study (see Appendices 7.4.11 and 7.5.3). Alternatively, tornado diagrams (discussed in Appendices 7.4.17 and 7.5.6) can also be used to visualise uncertainty in a deterministic model as a sensitivity analysis, or it could also be used to visualise the correlations between parameter attributes and the overall scores as shown in Figure 3.16 (10).

Figure 3.16 Tornado diagram of correlations between model attributes and overall score from stochastic model {Bujkiewicz, 2011}



Part 3.4 Discussion

Throughout the visualisation methodology process, we have adopted a systematic and structured approach to determining and creating visual displays. Although this may be seen as thorough and "academic", it is very time-consuming. Although it would be preferable to compare the ability to effectively communicate a benefit-risk message using visuals produced in this way to visuals produced simply on instinct, it is likely to be very difficult due to various unknown factors which may bias the results. We have to some extent illustrated in the graphical review and appraisal that many issues have been raised when visuals are produced on instinct – some may conflict with the good practice guidelines – and would benefit from some improvements. We did not experience as many critical comments in Phase II of systematically producing these visuals, but maybe it is because the issues have been addressed during the first phase.

So far, we have only incorporated our judgements and preference based on the combinations of the team's personal experience in the area of benefit-risk visualisations. The graphics created assume audience with low technical expertise in benefit-risk concepts and visualisations from our point of view. However, we feel very strongly that the choice of graphics to communicate benefit-risk to various stakeholders should actually be tested by the stakeholders themselves to ensure that the key message is fully understood. The subject is beyond the scope of this case study but need to be explored more in future research.

A recurring theme in this case study is the use of interactive and dynamic visualisations. We found interactive visualisation very appealing for use in benefit-risk assessment communications because of its flexibility to incorporate many features a static visual display could not. It may still be necessary to communicate benefit-risk messages using multiple graphs, but the use of 'dashboards' to display multiple graphs to allow single-screen view of the graphs is found to be desirable (11). It is also sometimes necessary to use multiple dashboards which are linked together, which would give the impression of having fewer graphs and having more control as to which information is displayed. It would also avoid overloading users with information by giving them time to digest the message layer by layer. In this case study we have not considered dashboards as an output but we envisage that it is likely to be the way forward in the arena of benefit-risk visualisation.

The critical decision to be made when generating interactive visualisations is which software package to use. There are various software vendors offering many graphical features at various prices (see PROTECT Visual Representation Review Report). Open-source software packages (free to use under GNU licence agreement) have made the spotlight with many being used for "infographics" in mass media such as the Guardian data blog (http://www.guardian.co.uk/news/datablog), the BBC news (http://news.bbc.co.uk/1/hi/technology/8562801.stm), art exhibitions (http://similardiversity.net) and television (http://vis.berkeley.edu/files/vizster.numb3rs.wmv).

We have chosen to reproduce the visuals in Wave 1 and generate interactive version of them using Tableau Public software because of its simplicity, and being free to use. We are generally satisfied with the quality of graphics it produces and the ease of use, but since data are uploaded and stored on external server, data confidentiality may be an issue. We have dealt with this by blinding the drug names. We also found that the graphics transition is not as smooth as we'd expected. Overall, it is still a good software package that is gaining momentum.

We acknowledge that interactive visualisation is not the solution or excuse to not presenting data clearly in the first place, but the technology opens up new avenues for the communication of benefit-risk assessment of medicines. Finally, whichever visualisation method is used – interactive or static – the data and analysis model used are far more important and should still be given more attention. The following Section 4 and Section 5 discuss the underlying data and statistical model that underpin the visual representations.

Section 4 Probabilistic uncertainty

Part 4.1 Introduction

4.1.1 Background – parameter uncertainty

The data on treatment safety and efficacy are derived from published studies which, even if carried out to the highest standards, are based on relatively small numbers of patients. The parameter values thus obtained are subject to sampling error, meaning there is a degree of uncertainty associated with each parameter. This is usually communicated in the literature by means of a 95% confidence interval.

Due to this uncertainty in the underlying parameter values, the overall benefit-risk balance (itself a function of the parameters) must also be uncertain. The Wave 1 case study dealt with the issue of uncertainty in a deterministic manner, by making a series of pre-determined perturbations to the central estimates of each parameter and observing the effect on the overall benefit-risk balance. This approach is often referred to as a sensitivity analysis. The perturbations were based on the quantiles of an assumed distribution for each parameter, which was approximately derived from the 95% confidence intervals reported in the source studies. This approach, although a reasonable starting point for investigating uncertainty, has a number of drawbacks:

- a deterministic, quantile-based approach does not fully reflect the shape of each parameter's distribution;
- the method does not provide a clear picture of the overall distribution of benefit-risk - only the variability relating to each individual parameter; and
- statistically the method lacks rigour - in particular, it conflates the frequentist and Bayesian schools of statistical thinking.

To expand on the last point: frequentist statistics holds that the true value of a parameter is a fixed, albeit unknown, quantity. This means that probabilistic statements regarding the value of a parameter are meaningless. However, it is possible to express a degree of confidence in an estimated parameter value in terms of likelihood; i.e. the probability of observing the results of a study conditional on the estimated parameter value. This is the concept underlying a 95% confidence interval, which is a range of parameter values that would be expected to replicate the observed results at least 5 times out of 100 were the study to be repeated many times.

Bayesian statistics, by contrast, regards a parameter as a random variable, following a probability distribution that represents our beliefs about the plausible range of values. It also provides a framework for updating the distribution to reflect the data observed in real-world studies. The resulting distribution, conditional on the observed data, is known as a posterior distribution. In the Bayesian world it is perfectly natural, using the posterior distribution, to make probabilistic statements about parameter values. This makes the Bayesian framework convenient for our purposes.

The 95% confidence interval and the (central 95% quantile of the) posterior distribution are often misinterpreted as being entirely analogous, and indeed they can be shown, in simple situations, to coincide. However, they are based on fundamentally different concepts. Interpreting a confidence interval from the literature in terms of a posterior distribution was a convenient approximation for the purpose of the Wave 1 case study, but for Wave 2 a more statistically robust method would be preferred.

4.1.2 Wave 2 objectives

The following objectives were identified for the Wave 2 case study:

1. Extend the uncertainty analysis from Wave 1 using a fully stochastic model

Instead of the deterministic sensitivity analysis employed in wave 1, a fully stochastic model will be used. Each parameter is to be assigned a probability distribution. Wherever possible, this will be the posterior distribution resulting from a Bayesian analysis of clinical data (but if this is not possible, the distributions derived in Wave 1 could still be used).

The distribution of the overall benefit-risk balance for each alternative, which is a mathematical function of the parameter values, can then be obtained by Monte Carlo simulation.

2. Compare the distributions of benefit-risk

The benefit-risk estimates in Wave 1 were point estimates, and it was straightforward to identify the treatment with the highest overall benefit-risk balance. With a distribution of overall benefit-risk balance for each treatment, the decision may not be as clear-cut. Methods for choosing between treatments based on their distributions will be investigated as part of the Wave 2 case study.

This is closely related to the question of how to present the results of a probabilistic benefit-risk analysis. This will be considered further in the Visual Review section.

Part 4.2 Methods

4.2.1 Statistical model for outcomes

The posterior distribution has two components: prior and likelihood. The prior distribution reflects any pre-existing knowledge or beliefs about the parameter. It may be that little or nothing about the parameter value is known in advance, in which case a non-informative prior may be used. The likelihood is the probability of observing the data conditional on a particular parameter value, and is analogous to the probabilities in a frequentist analysis. Mathematically, the posterior distribution is proportional to the product of the prior and the likelihood.

$$posterior \propto prior \times likelihood$$

It is possible to obtain a posterior distribution of each outcome, expressed on an absolute scale, in each arm (treatment or placebo) of a clinical trial by specifying:

- a *prior* distribution for the underlying value in that arm; and
- the *likelihood,* conditional on the underlying parameter value, of observing the particular outcome in that arm.

The prior distributions have been chosen because they form conjugate pairs with the likelihood: this means that the posterior can be expressed algebraically in the same form as the prior (but with parameters modified to reflect the observed data). A conjugate prior is convenient as it allows the posterior distribution to be neatly expressed in closed form. This facilitates the simulation stage of the analysis, as sampling from closed-form distributions is simple to implement in most statistical software packages. Being able to express a distribution in closed form also aids with validation and communication of the results. Other priors could however be used if considered appropriate, provided one has software capable of sampling from an arbitrary posterior distribution.

Although the aim for this case study was to select priors that are not unduly informative, the capacity of a Bayesian analysis to incorporate pre-existing information via the prior should not be forgotten.

4.2.2 Dichotomous outcomes

The majority of the outcomes in the value tree (PML, reactivation of serious viral infections, seizures, congenital anomalies, hypersensitivity reactions, flu-like reactions, elevation of transaminases, infusion/injection reactions, disability progression) are dichotomous at the patient level: a patient is recorded as either having experienced or avoided the event. The parameter we are interested in is the underlying probability π of each patient experiencing the event over a 2-year time period.

If n is the number of patients in a particular arm of a 2-year trial, and x is the number of events observed in that arm, then x can be assumed to follow a binomial distribution with parameters π and n. This gives the likelihood of the observed clinical data. A beta distribution with parameters a and b is specified as the prior distribution for π. By a well-known result, the posterior distribution of π is also beta:

$$\pi \sim Beta(a + x, b + n - x)$$

The binomial likelihood implicitly assumes that all patients have equal exposure to the risk of an event. For disability progression, however, this is not the case: if an individual undergoes a disability progression partway through a trial, they are removed from the pool of susceptible individuals, and therefore contribute less than 2 years of exposure. Such a situation is more accurately represented by a survival model that makes explicit allowance for the timing of each patient's disability progression. However, this requires full data at the patient level, which is generally not available. We use the binomial model here as an approximation, and estimate x by multiplying the reported 2-year survival probability by the total number of patients (effectively this is equivalent to assuming a full 2 years of follow-up for each patient).

The values chosen for the prior parameters were a=1/3, b=1/3 as it has been suggested that these values result in a non-informative prior (Kerman 2011). The effect of varying the prior parameters would later be investigated.

4.2.3 Relapses

Each patient may experience any number of relapses during a clinical trial. The parameter we are interested in is the underlying rate of relapses, r. This can be thought of as the expected number of relapses per patient over a fixed time period - in this case two years, although we work with the more conventional one-year rate and rescale this later in the analysis.

If T is the total patient-years of follow-up in a particular arm of a trial, and y is the total number of relapses, then y can be assumed to follow a Poisson distribution with parameter r. This gives the likelihood of the observed clinical data. A gamma distribution with parameters c and d is specified as the prior distribution for r. By another well-known result, the posterior distribution of r is also gamma:

$$r \sim Gamma\ (c + y, d + T)$$

When the values of T and y are not published, they can usually be estimated from other reported quantities. For example, the average relapse rate y/T is usually reported, and this can be used to obtain y given T or vice versa.

The values chosen for the prior parameters were c=1/3, d=0, as it has been suggested that these values result in a non-informative prior (Kerman 2011). Again, the effect of varying the prior parameters would later be investigated.

4.2.4 Administration

As there is no uncertainty associated with the convenience of the treatments, this outcome is modelled deterministically.

4.2.5 Calibrating on a common placebo group

It is necessary to measure the benefit and risk outcomes on an absolute scale, as this is the scale on which values are elicited. However, the studies were performed in different patient populations and at different points in time, so comparing absolute effects between treatment arms directly will confound the effect of the drug and the effect of the patient population. It is common practice in biostatistics to assume that the relative treatment or safety effect is robust to changes in the patient population. We compare drugs by taking a common placebo population, and finding the relevant absolute outcome for a drug by multiplying the outcome in the common placebo group with the appropriate relative rate, i.e. the ratio of the rates in the treatment and placebo arms of the appropriate trial.

This is a simple form of network meta-analysis. The placebo group in the natalizumab trial is chosen as the common placebo group for most outcomes, as this seems appropriate to a benefit-risk analysis of natalizumab. However where there are missing data in the natalizumab trial, the common placebo group is chosen as one with the available data.

For example, the relapse rate for beta-interferon is calculated as

$$r_b \times \frac{r_t}{r_p}$$

where r_b is the rate in the common placebo group, r_t is the rate in the treatment arm of the Beta-interferon trial and r_p is the rate in the placebo arm of the Beta-interferon trial.

Outcomes expressed as proportions are re-expressed as odds for the calibration stage, to avoid the possibility of obtaining a proportion greater than 1. After calibration the result is converted back into a proportion, as this is the scale on which the value functions have been elicited.

Since PML was not observed in any clinical trials, a relative effect cannot be extracted for this outcome. Instead, absolute rates from observational studies must be used. This is justifiable since the rate in patients not taking natalizumab is practically zero, so there can be no significant confounding due to population heterogeneities and calibration is therefore unnecessary. Arguably the same reasoning applies to the other rare events that do not occur in any placebo groups (seizures, hypersensitivity reactions, congenital anomalies, reactivation of serious infections). An alternative method, whereby the rates in the treatment arms are used directly, without calibration, will also be tested for these outcomes.

4.2.6 Data

As in wave 1, the data is drawn principally from three trials, one for each treatment (12), (13), (14).

For some outcome/treatment combinations, no data of sufficient quality could be found. In these cases it was assumed either that the outcome distribution was the same as in the common placebo group, or (for rare events that were not reported) that no events occurred in the relevant trial.

4.2.7 Simulations

Since the absolute outcome in each arm of each trial has a known posterior distribution, it is possible to use Monte Carlo techniques to sample from the distribution of the calibrated outcomes, which are algebraic functions of the absolute outcomes, and from the overall benefit-risk balance, which itself is an algebraic function of the calibrated outcomes.

The Monte Carlo simulations were implemented in the WinBUGS software package, which is specifically designed for sampling from the posterior distribution of arbitrarily complex Bayesian models. Since the posterior distribution of each outcome is expressed in closed form, however, specialist Bayesian sampling techniques are not strictly required and a wide variety of alternative software packages could be used.

4.2.8 Comparing benefit-risk distributions

A number of different methods were considered for comparing the distribution of benefit-risk between treatments:

- Comparing the mean and/or median of the distributions: although this method ignores stochastic uncertainty of benefit-risk, it may be useful for validation of the model (i.e. to check that the result is consistent with the deterministic approach in Wave 1).

- Distribution plots and/or box plots of benefit-risk: this method presents a clear visual summary of the distribution of benefit-risk for each treatment, including placebo, but may not be an effective aid to decision-making if there is significant overlap between the treatments. Lay audiences may find distribution plots difficult to interpret.

- Distribution plots and/or box plots of benefit-risk relative to placebo: generally it is of interest to determine whether a treatment has a better overall benefit-risk balance than placebo, but the benefit-risk distribution of placebo itself may be of little interest. Therefore this method visually presents the distribution of each treatment's overall benefit-risk balance relative to placebo. It is also possible to use an alternative treatment as the comparator, instead of placebo.

- Probability of treatment outperforming placebo: the proportion of simulations in which a particular treatment's overall benefit-risk was greater than placebo.

- Probability of treatment outperforming all others: the proportion of simulations in which a particular treatment had the highest overall benefit-risk.

Part 4.3 Results

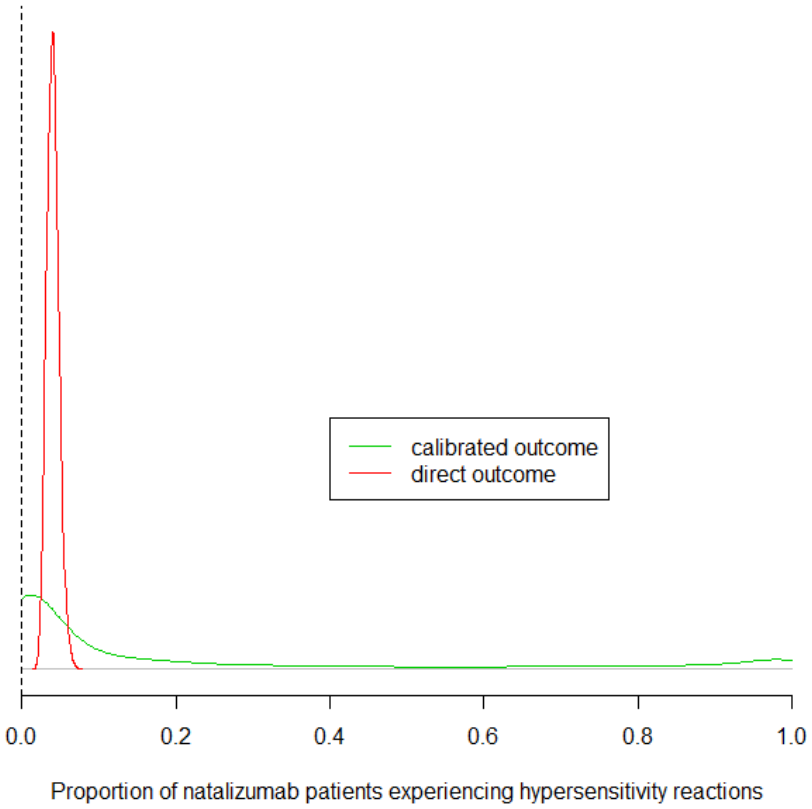
4.3.1 Validation

10000 simulations were run to obtain samples from the joint posterior distribution. The Appendix shows summary statistics for the posterior distributions of the clinical outcomes.

The medians of the benefit-risk distributions were all within 1% of the benefit-risk estimates obtained by deterministic MCDA (using the central estimates of the relevant clinical parameters).

The calibrated outcome distributions for rare events, although centred around the maximum likelihood estimates as one would expect, seemed unexpectedly wide: to give a particularly stark example, the 95% credibility interval for the proportion of natalizumab-treated patients experiencing hypersensitivity reactions was (0.000003, 0.9973). Inspection of a distribution plot reveals that most of the distribution is clustered near the maximum likelihood estimate of 4%, but with a very long, flat right tail and a slight secondary peak near 1, which seems likely to be an artefact of the beta prior. The model without calibration of rare outcomes (which instead compares the absolute outcome in the treatment arms directly) results in a more familiar, vaguely bell-shaped distribution plot for these outcomes (Figure 4.1).

Figure 4.1 Model output: Distribution of hypersensitivity reactions in patients taking natalizumab. The vertical axis indicates probability density.



As the distribution from the direct model is more precise and (arguably) appears more similar to a "typical" probability distribution, it is tempting to assume that this is the "correct" model for rare events. One should not be too hasty, however, in rejecting the calibrated model: one could argue that a long right tail is to some extent appropriate for rare events, as the paucity of observations means that estimates of the underlying rate are inherently uncertain. It may be simply our intuition – rather than anything contained in the data – that tells us these outcomes should always be "rare".

Notwithstanding this, it is difficult to deny that the calibrated distribution has an excessive amount of probability mass close to the value 1. Two methods dealing with this problem will later be discussed: the use of an alternative prior or a model with joint dependency between placebo and treatment arms in each trial.

Hereinafter the results from both models will be presented: the "indirect model" (where rare outcomes are calibrated) and the "direct model" (where the absolute values of rare outcomes are used directly).

4.3.2 Benefit-risk distributions

Figures 4.2(a) and 4.2(b) show density plots of the distribution of the overall benefit-risk score for each treatment in the direct and indirect model respectively. Note the long left tail in the benefit-risk distributions in the indirect model, which corresponds to the long right tail of the calibrated rare outcome distributions.

Figures 4.3(a) and 4.3(b) show the same distributions, this time represented by box-whisker plots. The "boxes" indicate the interquartile range of the distribution and the vertical "whiskers" extend to the minimum and maximum simulated values. The median of each distribution is indicated by a black diamond.

From a visual inspection of the distribution plots and box-whisker plots, it seems clear that natalizumab is the best treatment option. The bulk of the distribution lies to the right of (figure 4.3.2) or above (figure 4.3.3) the other treatments, corresponding to a higher overall benefit-risk balance. Choosing between the remaining three treatments is, however, not straightforward on the basis of these plots. Looking at the median of the distributions, Glatiramer acetate appears highest, followed by Beta-interferon, then placebo, but the difference is not great, and there are other aspects of the distributions to consider: for example, although placebo has a lower median benefit-risk than Glatiramer acetate, the latter is more likely to have a benefit-risk score below a particular cut-off point (corresponding to overall utility of around 0.92).

Figure 4.2 Overall benefit-risk distribution plot: direct model. The vertical axis indicates probability density.

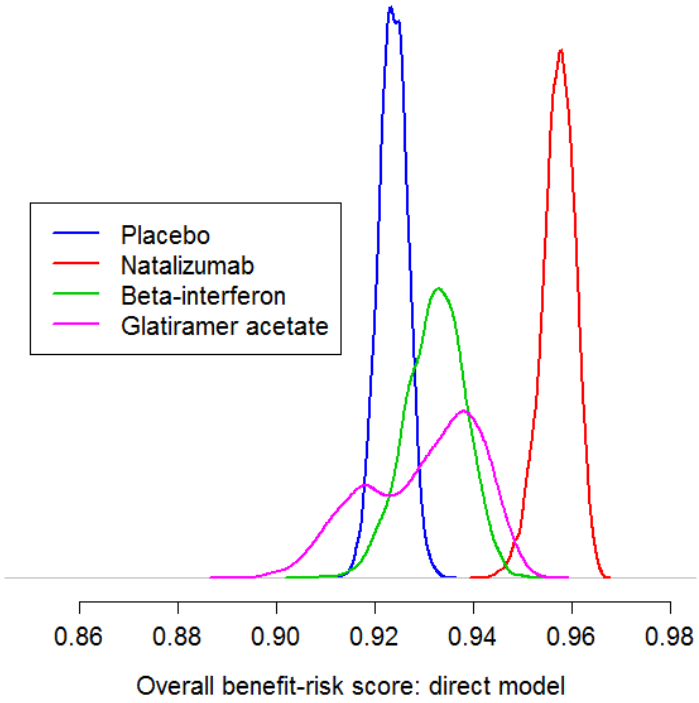


Figure 4.2(b) Overall benefit-risk distribution plot: indirect model. The vertical axis indicates probability density.

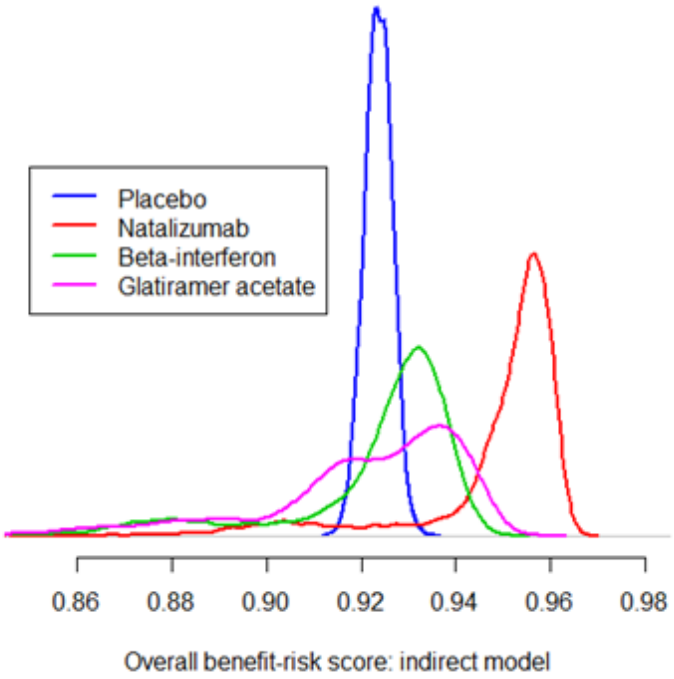
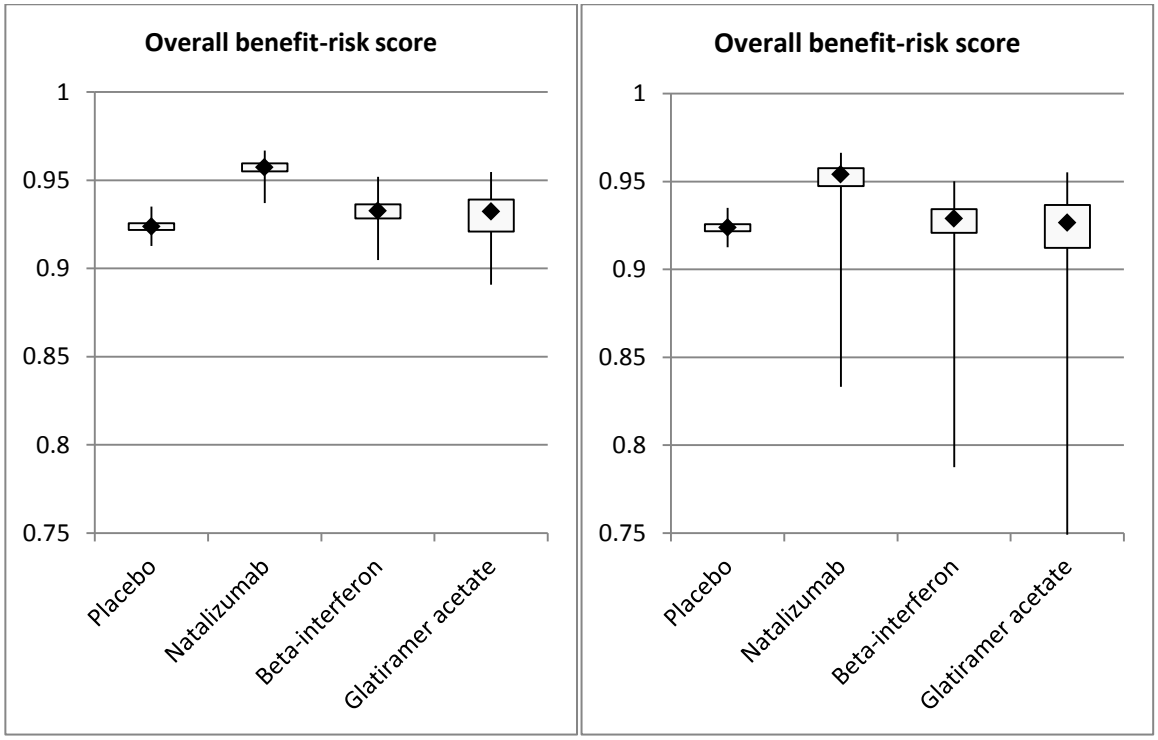


Figure 4.3(a) and 4.3(b) Overall benefit-risk box-whisker plot: direct model (a, left) and indirect model (b, right). The quantity on the vertical axis is the overall benefit-risk score, i.e. the overall utility of all 11 outcomes with the appropriate preference weights.



The medians and 95% credibility intervals of the benefit-risk distributions underlying figures 4.2 and 4.3 are shown in Table 5 below.

Table 5 Medians and 95% credibility intervals for overall benefit-risk score.

TREATMENT	OVERALL BENEFIT-RISK: MEDIAN (95% CI)	
	Direct model	Indirect model
Placebo	0.924 (0.918, 0.929)	0.924 (0.918, 0.929)
Natalizumab	0.958 (0.950, 0.963)	0.954 (0.897, 0.962)
Beta-interferon	0.933 (0.920, 0.943)	0.929 (0.869, 0.942)
Glatiramer acetate	0.932 (0.907, 0.947)	0.926 (0.864, 0.946)

4.3.1 Benefit-risk distributions relative to placebo

Figures 4.4(a) and 4.4(b) show density plots of the distribution of the overall benefit-risk score for each treatment relative to placebo. Figures 4.5(a) and 4.5(b) show box-whisker plots representing the same distributions. The dashed line indicates where benefit-risk is equal to placebo.

Figure 4.4(a) Overall benefit-risk relative to placebo, distribution plot: direct model. The vertical axis indicates probability density.

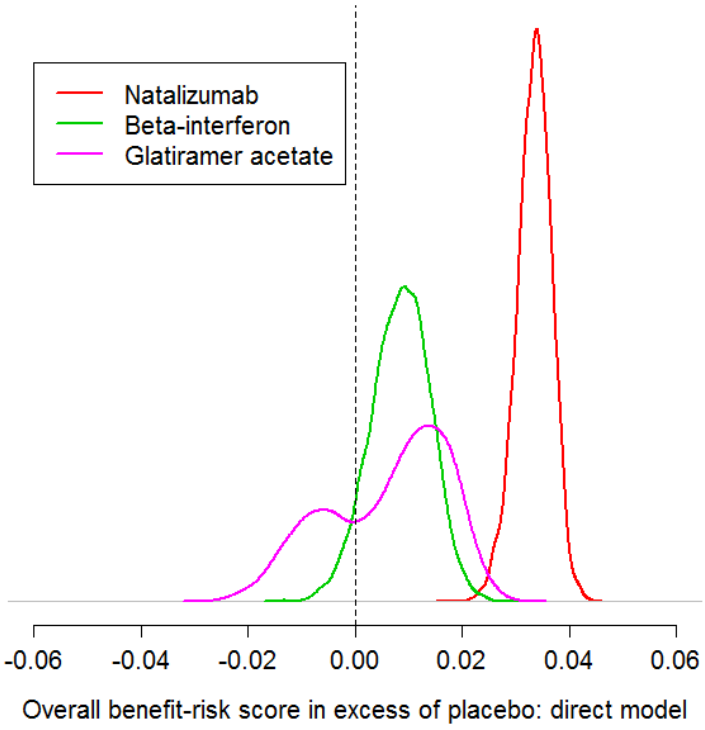


Figure 4.4(b) Overall benefit-risk relative to placebo, distribution plot: indirect model. The vertical axis indicates probability density.

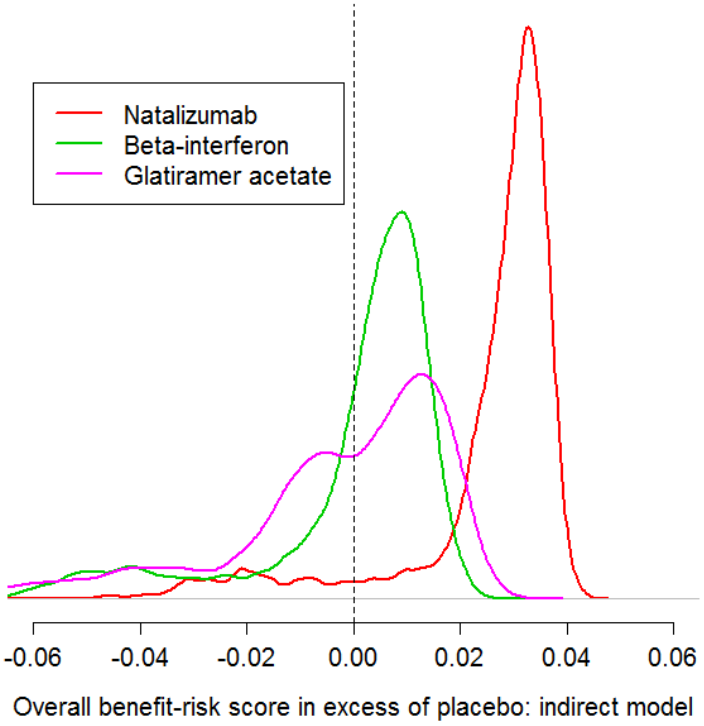
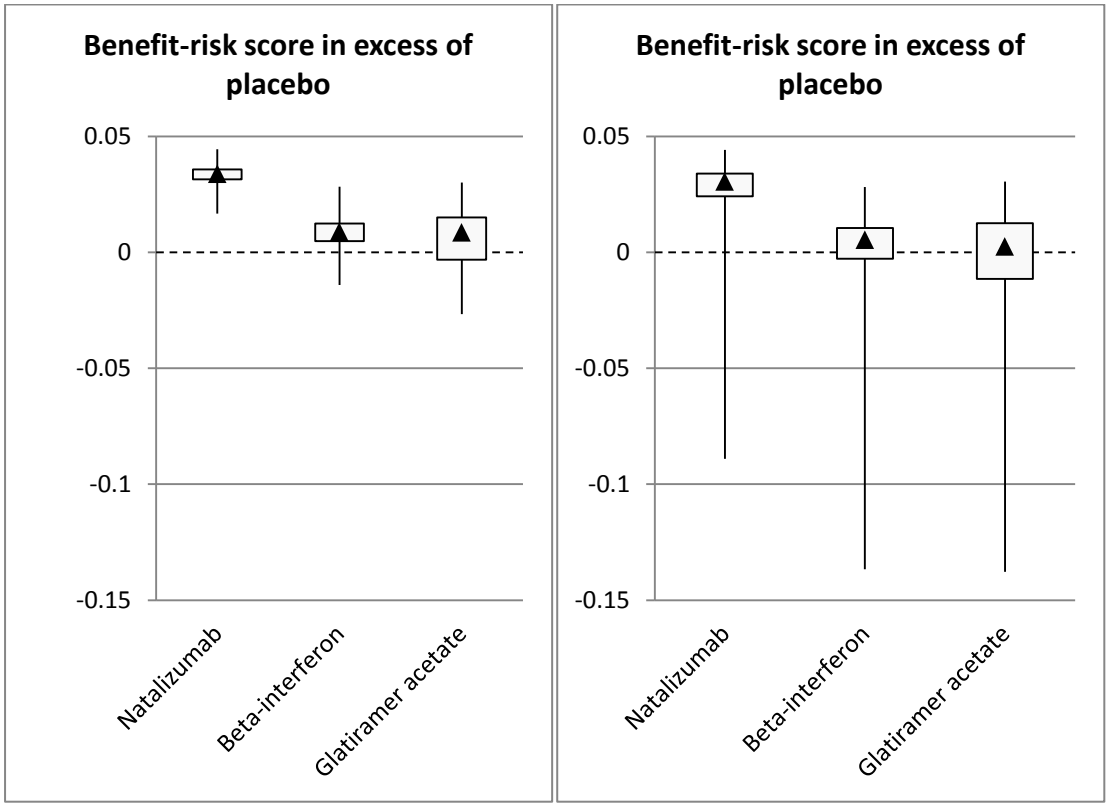


Figure 4.5(a) and 4.5(b). Overall benefit-risk relative to placebo, box-whisker plot: direct model (a, left) and indirect model (b, right). The quantity on the vertical axis is the overall benefit-risk score in excess of placebo.



The medians and 95% credibility intervals of the benefit-risk distributions relative to placebo underlying figures 4.4 and 4.5 are shown in Table 6 below.

Table 6 Medians and 95% credibility intervals for overall benefit-risk score in excess of placebo.

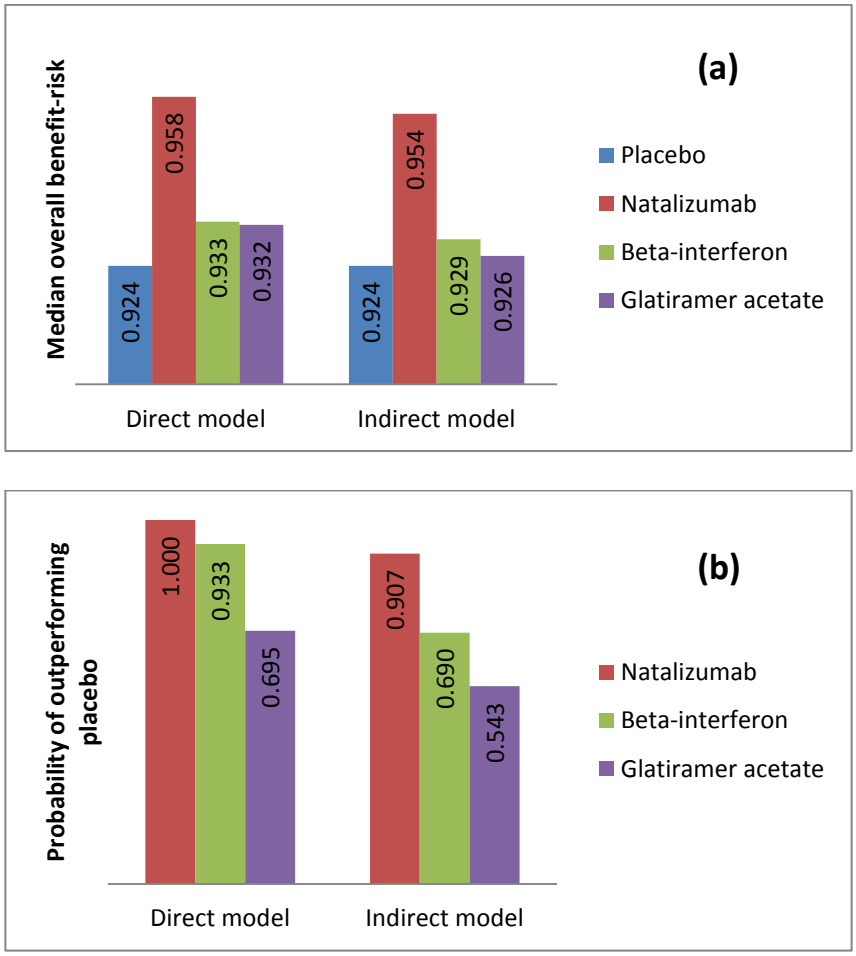
TREATMENT	BENEFIT-RISK IN EXCESS OF PLACEBO: MEDIAN (95% CI)	
	Direct model	Indirect model
Natalizumab	0.034 (0.026, 0.039)	0.030 (-0.027, 0.039)
Beta-interferon	0.009 (-0.003, 0.019)	0.005 (-0.054, 0.018)
Glatiramer acetate	0.009 (-0.016, 0.023)	0.002 (-0.060, 0.022)

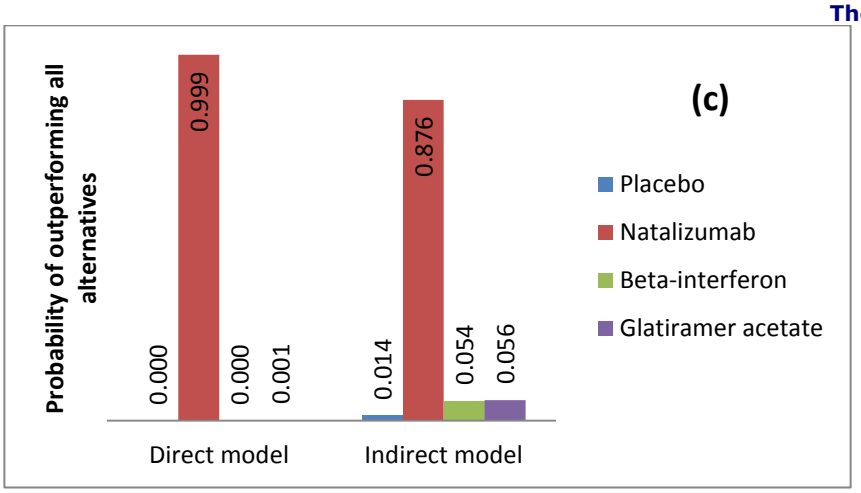
4.3.2 Metrics for comparing treatments

Figure 4.6 (a-c) compares the treatments based on the following measures:

(a) median overall benefit-risk score;
(b) probability of outperforming placebo; and
(c) probability of outperforming all alternatives.

Figure 4.6 Comparison of treatments according to three different measures: (a) median overall benefit-risk score; (b) probability of outperforming placebo; (c) probability of outperforming all alternatives.





Natalizumab is the optimal choice under all three metrics (and in both models). Placebo is consistently ranked as the worst alternative - although it does not appear in figure 4.3.6(b) since it does not make sense to compare placebo with itself, it is worth noting that all drugs had at least an even (50%) chance of outperforming placebo, implying that placebo is the worst option.

It is interesting to contrast how the three methods distinguish between beta-interferon and glatiramer acetate. Beta-interferon has only a very slight advantage in terms of median benefit-risk (Figure 4.6a), but it outperforms placebo significantly more often (Figure 4.6b). However, glatiramer acetate is the top-ranked treatment slightly more frequently than beta-interferon (Figure 4.6c). These results may seem counterintuitive, but they can be traced back to the relatively high variance and skewness of the benefit-risk distribution for glatiramer acetate (Figure 4.2).

4.3.3 Sensitivity to choice of prior

The impact of the choice of prior was investigated by repeating the analysis with a number of different conjugate beta and gamma priors for the outcome variables (beta for dichotomous events, gamma for relapses).

- *Priors expected to be non-informative: Beta (1/3, 1/3), Beta (0.5, 0.5), Beta (1, 1), Gamma (1/3, 0), Gamma (0.5,0), Gamma (1,0)*

  Changing between these priors did not have a significant impact on the mean or median of the posterior distributions of the outcomes, which all remained very close to the maximum likelihood estimate based on the data, although the width of the distributions was affected, with some becoming wider and some narrower. The overall benefit-risk distributions were not significantly affected in the direct model. In the indirect model with calibrated rare events, the overall benefit-risk distributions were narrower when the beta parameters were higher.

  The order in which the drugs were ranked (based on the three metrics already discussed) was not affected by the choice of prior, except in one case: the median overall benefit-risk for Glatiramer acetate slightly exceeded that of Beta-interferon when using a beta (1,1) prior.

  Under all of these priors, the calibrated outcome distributions for rare events exhibited a slight peak near the value 1, as seen in Figure 2. This may be an artefact that is common to all beta priors. Since the methodology adopted for this analysis was dependent on the beta-binomial conjugate relationship, however, the use of alternative forms for the prior distribution could not easily be investigated.

- *Informative priors for rare events: Beta (1/3, N + 1/3)*

As noted above, the calibrated outcome distributions for rare events were wider than expected: one way that was considered to address this was to use an informative prior to convey the rareness of an event. This can be achieved in the beta-binomial model by adding a large number N to the second beta parameter, which is equivalent to having observed zero events out of a notional population of N patients (in addition to the clinical trial data).

This approach succeeded in reducing the tail size of the calibrated outcome distributions, but as N became very large the impact of the trial data was reduced almost to zero, with the posterior distributions having been "shrunk" towards the prior. This meant that there was effectively no observable difference between the treatments for these outcomes, and this sometimes had a significant impact on how the drugs were ranked. The validity of this approach, therefore, depends heavily on the choice of N, which may be difficult to objectively justify unless based on hard data.

Part 4.4 Discussion

4.4.1 Impact on benefit-risk analysis

Carrying out a probabilistic benefit-risk analysis has the following advantages over the simpler deterministic approach:

- *Providing a sense of the significance of differences between treatments*: in deterministic MCDA based on point estimates, as there is no measure of uncertainty, one cannot gauge the statistical significance of the difference in benefit-risk score. It is therefore impossible to discount the possibility that any observed differences are simply due to chance. For example, in the case study, Beta-interferon outperforms Glatiramer acetate on the basis of a deterministic analysis. It is only when uncertainty is introduced that one gets a sense of just how flimsy this result is. Similarly, it is only in the light of the stochastic analysis that the significance of natalizumab's advantage over the other treatments becomes clear.

- *Demonstrating the robustness of a decision to data uncertainty*: in the case study, figure 7c shows that natalizumab is the optimal treatment for the vast majority of plausible data values, improving confidence in the results of the deterministic analysis.

- *Helping to communicate uncertainty*: decision-makers may place excessive confidence in results that are presented to them deterministically, even if they are reminded of the underlying variability. It is harder to ignore uncertainty if the results are presented as distributions or probabilities. Furthermore, a statement regarding the probability of an uncertain outcome is likely to be easier for a layperson to interpret than a sensitivity analysis.

- *Highlighting the need for further evidence-gathering*: if a posterior distribution for a parameter of interest is judged to be excessively wide, it may indicate the need for further studies to increase the available evidence base.

Sensitivity analysis goes some way towards addressing some of these points, but suffers from the drawbacks discussed in Section 4.1.1.

Another advantage of Bayesian modelling in particular is the capacity to update the analysis in the light of new clinical evidence. This is straightforward in the Bayesian framework: the posterior distributions of the outcomes from the first analysis are used as priors in the second. This gives a natural and robust method for combining the old and new evidence.

The drawback of this approach compared to a deterministic analysis is its relative complexity. Each decision situation, with its own set of criteria and data, requires a bespoke stochastic model. A certain amount of statistical expertise is therefore required, and the validity of the analysis may be called into question if the model structure is judged to be flawed. Implementing an analysis based on probabilistic simulations may also be beyond the technical capabilities of many decision-makers.

4.4.2 Similar methods developed elsewhere

There have been other attempts to use stochastic models to allow for parameter uncertainty in benefit-risk. Bayesian modelling has been employed to derive distributions of benefit-risk in relation to warfarin, an anticoagulant (15) . Expressions have been derived for the variability of benefit-risk (16) . However, these results were based on a measure of benefit-risk known as net clinical benefit (NCB), which corresponds to a special case of MCDA in which only dichotomous outcomes and a particular form of value function are permitted, greatly simplifying the mathematics involved. The approach presented here can be applied more widely within the MCDA framework.

Stochastic multi-criteria acceptability analysis (SMAA) is an extension of MCDA, designed for situations where weights have not been elicited and therefore remain unknown (17) . SMAA ranks alternatives using one of the metrics used here (the proportion of simulations in which each drug outperforms all others), but allows for uncertainty in preference weights as well as clinical parameters. This method has been applied to medical benefit-risk (18) , using specialised SMAA software to carry out a benefit-risk analysis allowing for uncertainty in both clinical parameters and preferences (19). In the context of IMI-PROTECT, however, it is more appropriate to elicit preferences in advance, as it is important that stakeholders' views are reflected in the benefit-risk analysis.

4.4.3 Comparison of visualisations

The distribution plots of overall benefit-risk (Figure 4.2) are arguably the most informative visualisations from a statistician's point of view, but they may be difficult for a lay audience to interpret. This is an important consideration if decisions are to be made at the patient level. The box-whisker plots may be somewhat easier to explain, but do not convey as much information regarding the shape of the distributions. Moreover, it is not easy on the basis of any of these plots to decide between distributions that overlap significantly, such as placebo and Glatiramer acetate.

Such comparisons are made easier by examining the distribution of one alternative relative to another, such as in figures 4.3.4 and 4.3.5, where the treatments are benchmarked against placebo. It is easier to infer from these plots that Glatiramer acetate outperforms placebo, particularly in the direct model.

It is worth mentioning that it is also possible to examine the distribution of treatments' overall benefit-risk relative to each other, rather than placebo, if this is of particular interest. This might be the case, for example, if a healthcare provider with limited funds is forced to choose between two competing treatments. To avoid presenting an overwhelming quantity of figures and charts, this has not been illustrated here, but it is analogous to the placebo case.

All the distribution plots and box plots show the underlying utility scale on which the results are measured, making it possible to gauge the clinical significance of differences between the treatments. This detail cannot be extracted from the probability metrics, which only give information on the order in which the treatments are ranked.

The probability of outperforming placebo was effective at discriminating between beta-interferon and glatiramer acetate, which performed almost equally well under the other metrics. However, there is no way of knowing by what extent each treatment outperformed placebo; the difference may be clinically insignificant. This could be overcome by calculating how often each treatment outperforms placebo by a pre-determined, clinically important difference. Determining the appropriate size of this margin, however, may not be without its own difficulties. Even

with such a margin, this method would fail to distinguish between, say, a drug that always performs twice as well as placebo and one that always performs three times as well, even though it is clear which is better.

The last metric, the probability of outperforming all alternatives, would have no trouble making this distinction, although it too cannot determine whether differences are clinically significant unless an explicit margin is incorporated. It is striking to note how clearly natalizumab dominates on the basis of this metric: one consequence of this, however, is that there is little discrimination between the alternative treatments (this is particularly stark in the direct model results). This metric is probably the easiest to interpret and communicate – it simply assigns each treatment a probability of being the optimal choice.

In summary, then, there are advantages and disadvantages to all of the visualisations and metrics that were used, and the choice of which is appropriate in a particular situation depends heavily on the target audience, the nature of the distributions being compared, and the relative importance of clinical and statistical significance.

4.4.4 Validity of statistical model

It seems likely that the wide distributions in the "indirect model" arise because the values of an outcome in the placebo and treatment arm of the same trial are assumed to be independent. This means that in a particular simulation there is a reasonable chance of observing a very low rate in the placebo arm and a very high rate in the treatment arm or vice versa, resulting in an unstable estimate of the relative effect of treatment. On reflection, this would seem to contradict the assumption that relative effects are robust to changes in population. In reality, if this assumption holds, we would expect to see correlation between the outcomes in the placebo and treatment arm.

This difficulty could perhaps be overcome with a more sophisticated model that specifies a joint likelihood for each outcome in both placebo and treatment arms simultaneously. One disadvantage of this approach, however, is that it would require an assumption regarding the precise nature of the interdependence between the two arms, and it may not be obvious what form the relationship should take. It is worth noting that this is only a significant issue for rare events, since when the numbers are small, a relatively minor change in the outcome in either arm leads to a large swing in the relative effect of treatment. Furthermore, any such changes to the model structure may not have any impact on the ranking of the treatments, since all treatments would be affected more or less equally.

The approach taken to combining the results of the different trials relies on the assumption that the relative effect of treatment is robust to changes in the characteristics of the population such as age, socioeconomic status, other therapies being used, and so forth. This assumption, although commonplace in biostatistics, may not always hold.

Another aspect of the model that may be criticised as insufficiently realistic is that it assumes the various outcomes occurring in a clinical trial are independent from one another. In reality one might expect the outcomes to be interdependent in some way – for example, the risk of relapse and disability progression may be correlated. A model that allowed for such interdependencies would, however, require data either from a vast number of trials or at the individual patient level, neither of which were available for this case study.

4.4.5 Choice of prior

In general, the choosing of different priors (within reasonable limits) does not affect the drug rankings, because changing prior affects all treatments more or less equally. In a decision-making context, where the relative performance of treatments is of principal importance, the overall benefit-risk score is effectively an arbitrary measure designed for making comparisons rather than a meaningful absolute measure. The choice of prior may therefore be of lesser importance than in applications requiring accurate outputs on an absolute scale, as long as there is consistency between the alternative treatments.

Nevertheless, it is recommended that care is taken to select an appropriate prior, for the following reasons in particular:

- excessive shrinkage towards the prior may swamp differences in the data, reducing the model's power to differentiate between treatments;
- in certain situations the choice of prior can in fact affect the drug rankings (this was observed for the beta (1,1) prior);
- checking whether the outcome distributions appear sensible given one's knowledge of the data is a useful model validation tool and an aid to communicating the credibility of the results; and
- it may be desirable for the posterior distributions of clinical outcomes obtained from the model to be made available for other uses.

Part 4.5 Acknowledgements

Some of the work relating this section on probabilistic uncertainty was written up separately by Ed Waddingham and submitted to Imperial College London to meet the requirements of an MSc in Modern Epidemiology. There is some duplication of text between the MSc thesis and this document.

Section 5 Discussion

Part 5.1 Limitations

5.1.1 Time horizon of the decision problem

A notable limitation is the fixed time horizon of the benefits and risks. We are using a two-year time horizon largely because this is the length of the pivotal trials for natalizumab and its comparators, and so the bulk of the available data on benefits and risks relates to a two-year period. In principle, however, any time horizon could be used to frame the decision problem, and the choice of the appropriate horizon should fall to the decision-maker.

Furthermore, stakeholders' attitudes to benefits and risks are likely to be time-dependent. For example, a patient may be prepared to accept an elevated risk of death in a few years' time as long as their disabling symptoms are minimised for the next year or two. To allow for this, a BR method that reflects the changing nature of benefits and risks over time (eg by incorporating a variable time horizon) would therefore be desirable. The interactive visualisations described in this report could be developed further to allow for a user-defined time horizon. This would be an interesting area for further research, but would require considerably more data than was available for the purpose of this case study.

5.1.2 Scope of probabilistic uncertainty

The overall benefit-risk score for each treatment is a function of:

- clinical parameters, estimated here from the results of published clinical studies; and
- preference parameters, estimated here from the results of the elicitation study carried out in Wave 1.

Both kinds of estimate may not accurately reflect the true average parameter value in the decision population, for reasons including:

- *sampling error* – whereby the average quantity observed in the study sample differs from the study population average due to random variability
- *heterogeneity between the study population and the decision population* –whereby the average value in the study population may not be the same as the average value in the decision population because the populations differ in some respect
- *study design factors* – parameter estimates from randomised clinical trials may be biased due to the effects of treatment crossover, withdrawal and loss to follow-up; observational studies are prone to additional problems such as confounding and reverse causality; preference elicitation studies may employ a number of designs that differ in various ways (such as the underlying assumptions and the handling of inconsistent judgements) and therefore give different estimates of the decision parameters
- *approximations made in the data extraction*– the source data may need to be transformed in some way (eg because the time horizon of the decision problem is not the same as that of the reference study) – usually this involves additional assumptions and/or approximations that may introduce further inaccuracy .

The probabilistic model described in section 4 only allowed for uncertainty due to sampling error. Furthermore, it only allowed for the sampling error of the clinical parameters; preference information was regarded as fixed.

It is also possible, although more complicated, to allow for uncertainty of preference data. This could be done, for example, by using elicitation experiments to construct preference distributions or by simulating all possible combinations of weights. The latter approach is employed in the technique known as SMAA. This approach was not followed in this case study, partly due to time constraints, but also because a priori elicitation of preference

information was one of the aims of the project. The preference elicitation experiments are, at the time of writing, still ongoing and the results will be reported in due course.

Part 5.2 Conclusions from the wave 2 case study

The catalogue of visualisations emerging from the visual review workstream was found to be useful, with visualisations that could be employed at all stages of BR assessment. From the point of view of the case study team, the visualisations were generally seen to be clear and appropriate, and the interactive visuals were particularly well received, but there should be further testing on external audiences.

The use of a probabilistic model can further illuminate comparisons between treatments based on their benefit-risk scores – in particular, it clearly shows how robust the comparisons are to uncertainty in the data. But a probabilistic model is difficult to set up and may only add value if the decision-maker's attitude to uncertainty is well understood. We expect that, where a decision is to be made between competing treatments, most decision-makers will continue to choose whichever treatment has the best benefit-risk score based on a deterministic analysis. However, being able to clearly visualise the distribution of benefit-risk does present many advantages, as set out in Section 4.4.1.

The two sub-tasks of this case study are not entirely self-contained; there are links between them

- Visualisations are a key part of understanding probability distributions (eg distribution plots, box-whisker plots).
- Incorporating uncertainty into visualisations can be difficult but should not be ignored; visualisations that are too simple may present the unintended message that there is no uncertainty.

Section 6 References

1. Rosati G. The prevalence of multiple sclerosis in the world: An update. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology JID - 100959175. 0417.

2. Compston AF, Coles A. Multiple sclerosis. Lancet JID - 2985213R. 0425.

3. GUIDELINE ON CLINICAL INVESTIGATION OF MEDICINAL PRODUCTS FOR THE TREATMENT OF MULTIPLE SCLEROSIS, CPMP/EWP/561/98 rev. 1. [Internet]. Available from: http://www.ema.europa.eu/ema/pages/includes/document/open_document jsp?webContentId=WC500003485.

4. Assessment report for NATALIZUMAB. EMA/H/C/000603/A20/0029. [Internet]. Available from: http://www ema europa eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000603/WC500095872 pdf.

5. Kappos LF, Bates DF, Edan GF, Eraksoy MF, Garcia-Merino AF, Grigoriadis NF, et al. Natalizumab treatment for multiple sclerosis: Updated recommendations for patient selection and monitoring. Lancet neurology JID - 101139309. 0922.

6. Transatlantic workshop on drug-related progressive multifocal leukoencephalopathy (PML): Workshop proceedings. EMA/674403/2011 rev. 1. [Internet]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/09/WC500111562.pdf.

7. European Medicines Agency. Third report on the progress of the interaction with patients' and consumers' organisations during 2009. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2010/11/WC500099257.pdf: 2010.

8. IMI-PROTECT WP5. Review of visual representation for benefit-risk assessment of medication
Stage 2 of 2. www.imi-protect.eu: 2013.

9. Wickens CD, Lee J, Liu YD, Gordon-Becker S. An introduction to human factors engineering. 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice-Hall; 2004.

10. Bujkiewicz S FAU - Jones, Hayley,E., FAU JH, FAU LM, FAU CN, Hawkins NF, Squires H FAU - Abrams, Keith,R., et al. Development of a transparent interactive decision interrogator to facilitate the decision-making process in health care. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research JID - 100883818. 1007.

11. Few S. Information dashboard design. 1st ed. Sebastopol, CA: O'Reilly Media, Inc.; 2006.

12. Polman C, O'Connor P, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. The New England journal of medicine JID - 0255562. 2006.

13. Jacobs L, Cookfair D, Rudick R, Herndon R, Richert J, Salazar A, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. the multiple sclerosis collaborative research group (MSCRG). Annals of neurology JID - 7707449. 1996.

14. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: Results of a phase III multicenter, double-blind placebo-controlled trial. the copolymer 1 multiple sclerosis study group. Neurology. 1995;45(7):1268.

15. Sutton AJ, Cooper NJ, Abrams KR, Lambert PC, Jones DR. A bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty. J Clin Epidemiol. 2005;58(1):26-40.

16. Sinclair JC, Cook RJ, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used?: Derivation of the threshold number needed to treat and the minimum event rate for treatment. J Clin Epidemiol. 2001 3;54(3):253-62.

17. Lahdelma R, Hokkanen J, Salminen P. SMAA - stochastic multiobjective acceptability analysis. Eur J Oper Res. 1998 4/1;106(1):137-43.

18. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med. 2011;30(12):1419-28.

19. Tervonen T. JSMAA: Open source software for SMAA computations. International Journal of Systems Science. 2012;(to appear).

Section 7 Appendix

Part 7.1 BRAT Step 1 - Define the decision context

The group discussed and agreed a proposal on objective and assumptions of the assessment and the contextual basis for it as outlined below:

Table 7 Decision context resulting from the discussion

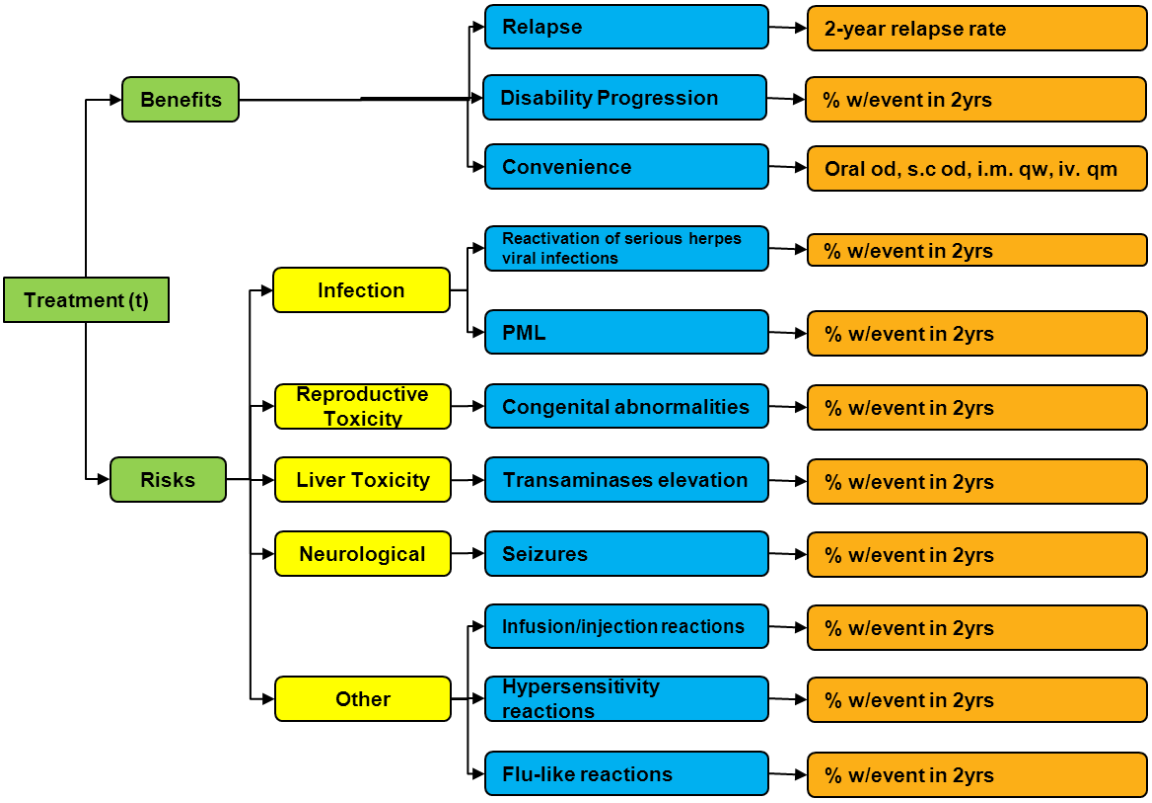
Objective	1. Should natalizumab be given marketing approval at the time of first registration? 2. Should natalizumab be kept on the market given that increased episodes of PML were observed?
Indication	Monotherapy for the treatment of highly active RRMS (relapsing remitting multiple sclerosis) SmPC section 4.1 indication in detail: NATALIZUMAB is indicated as single disease modifying therapy in highly active relapsing remitting multiple sclerosis for the following patient groups: Adult patients aged 18 years and over with high disease activity despite treatment with a beta-interferon. These patients may be defined as those who have failed to respond to a full and adequate course (normally at least one year of treatment) of beta-interferon. Patients should have had at least 1 relapse in the previous year while on therapy, and have at least 9 T2-hyperintense lesions in cranial Magnetic Resonance Image (MRI) or at least 1 Gadolinium-enhancing lesion. A "non-responder" could also be defined as a patient with an unchanged or increased relapse rate or ongoing severe relapses, as compared to the previous year. or Adult patients aged 18 years and over with rapidly evolving severe relapsing remitting multiple sclerosis defined by 2 or more disabling relapses in one year, and with 1 or more Gadolinium enhancing lesions on brain MRI or a significant increase in T2 lesion load as compared to a previous recent MRI
Drug	NATALIZUMAB (natalizumab)
Formulation/Dose	natalizumab 300mcg, iv, qm
Comparators	Beta-interferon (Interferon beta-1a): RMMS, 30mcg, im, qw Glatiramer acetate (Glatiramer acetate):RMMS, 20mg, sc, qd Placebo
Population	Patients with RRMS
Time Frame for Outcomes	24 months
Stakeholder	Decision maker Regulator (at EMA), taking the patient perspective. The regulator makes the decision, using the values and weights of a patient

Part 7.2 BRAT Step 2 - Identify benefit and risk outcomes

The value tree was prepared by some individual experts based on the available data. In general the Summary of Product Characteristics (SPC) for natalizumab and the European Assessment report by EMA were chosen as the foundation for this assessment as the data outlined could be considered as validated by Regulatory Authorities. A group decision was taken on the key benefits and risks to be included. Similarly it was agreed that PML should be

included at time of simulated initial approval assessment as PML could be considered as a potential risk identified from combination treatment at this point in time.

Figure 7.1 Value tree



Part 7.3 Gathering objective data: BRAT Step 3 - Identify and extract source data

This section details the sources of objective evidence this is used in the benefit-risk assessment. A two year time horizon for the decision was chosen as this is considered the minimum duration for the drugs in the scope to demonstrate efficacy. The scope is limited to Phase III studies as these have this time horizon and were thought to be able to provide the most reliable data and a sufficiently rich source of safety data for the analysis. The sources of evidence to extract relevant data from are:

- European public assessment reports (EPAR), “product information” and “scientific discussion”
- Literature search.

Periodic Safety Update Reports (PSURs) are not an option for wave 1 because of time restraints required to extract the relevant data.

7.3.1 Literature searching

Two databases were searched from English-language reports of controlled clinical trials assessing the efficacy and safety of alternative MS drugs for relapsing-remitting multiple sclerosis: EMBASE from 1980 to July 2011 and Ovid MEDLINE from 1948 to July 2011. In addition, reference searching from review papers and expert’s suggestion was also performed.

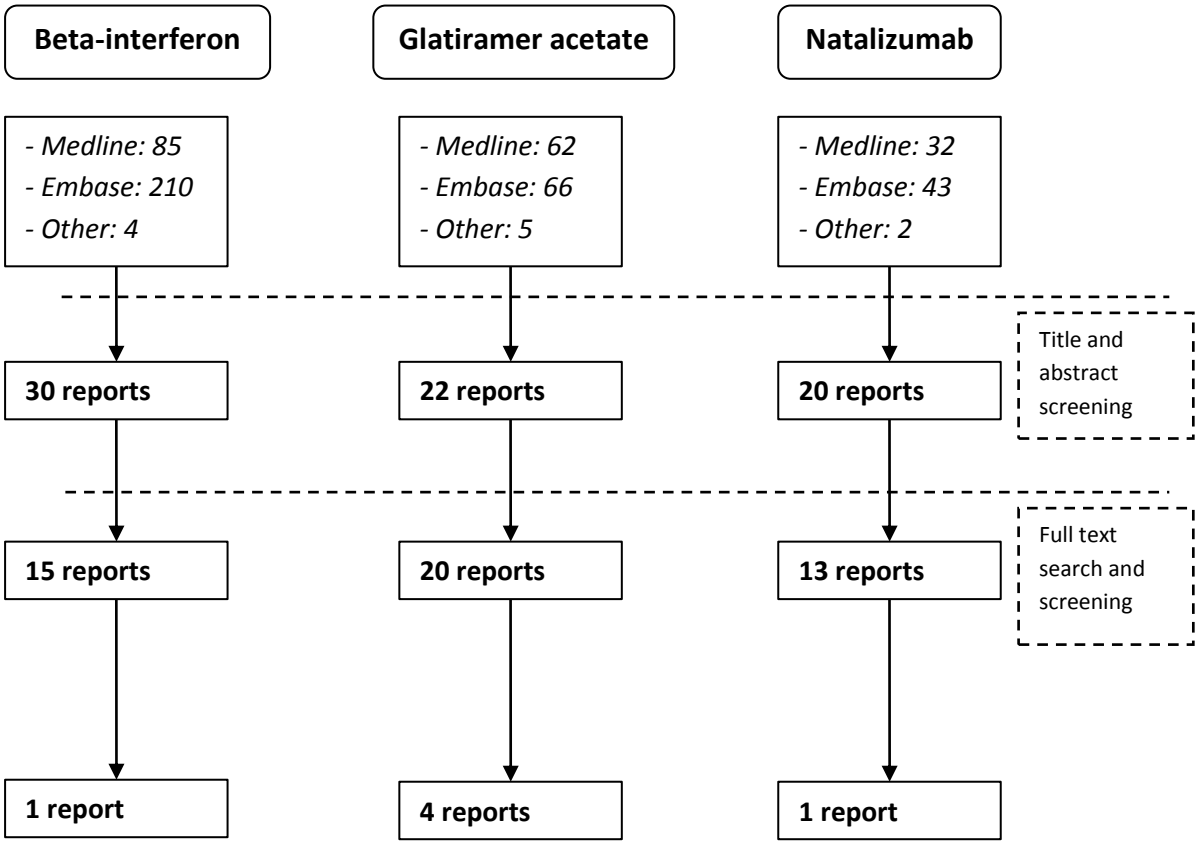
Database searching was performed for each drug using a similar search strategy, the details of these search terms is given in Table 9.

Table 8 Literature searching and data selection strategy

<table>
<tr><td colspan="2">LITERATURE SEARCHING AND DATA SELECTION</td></tr>
<tr><td rowspan="2">Interferon Beta-1a
Interferon beta.m_titl. (title)
Beta interferon.m_titl.
Avonex.m_titl.
[1] OR [2] OR [3]
Multiple sclerosis.mp. and multiple sclerosis/ (Map term to subject heading)
[4] AND [5]</td><td>Glatiramer Acetate
Glatirameracetate.m_titl. (title)
Copaxone.m_titl.
Copolymer 1.m_titl.
Cop 1.m_titl.
[1] OR [2] OR [3] OR [4]</td></tr>
<tr><td>Natalizumab
Natalizumab.m_titl. (title)
Tysabri.m_titl.
[1] OR [2]</td></tr>
<tr><td colspan="2">Additional limits : English language AND Human AND (RCT or controlled clinical trial or phase II clinical trial or phase III clinical trial or phase IV clinical trial)</td></tr>
<tr><td colspan="2">Inclusion criteria for title and abstract screening: Clinical Trials for Relapsing-Remitting Multiple Sclerosis</td></tr>
<tr><td colspan="2">Exclusion criteria for title and abstract screening: Other Interferon beta not Avonex</td></tr>
<tr><td colspan="2">Inclusion Criteria for primary input data in MCDA model: Randomised placebo-controlled phase III clinical trial</td></tr>
</table>

Results from literature searching and data collection for clinical trials are illustrated in Figure 7.2.

Figure 7.2 Literature searching results



At the final screening stage (full text screening), four reports of three Phase III placebo-controlled clinical trials have been found for glatiramer acetate while only one Phase III placebo-controlled trial has been found in each drugs for natalizumab and Beta-interferon.

After discussion with clinicians and statisticians, three reports of Glatiramer acetate were are excluded due to their methodology and quality. Rationale for trials exclusion at this stage is presented in Table 10.

Table 9 Excluded clinical trials

Trials	Rationale for exclusion
Bornstein 1987	Only pilot trial with small sample size and using match-pair randomisation
Johnson 1998	Extension period over 2 years from the core trial (Johnson KP 1995)
Comi 2001	Trial period was only 9 months

Details of the studies from which data are extracted are given in Table 11

Table 10 Details of studies used in the analysis

Drug	Author	Year	Exposure	Comparator	Details
Natalizumab	Polman CH	2006	28 Months	Placebo	RRMS. Nov 2001: 942 patients - multiple centres and countries. 856 patients (91%) completed the 120-week trial
Beta-interferon	Jacobs LD	1996	24 Months	Placebo	RR MS. Early 1993: 301 patients - 4 clinical centers US.
Glatiramer acetate	Johnson KP	1995	24 months	Placebo	RRMS. October 1991-May 1992: 251 randomised. 215 completed 2-year trial. US population.

Efficacy data were only extracted from publications whereas safety data were also extracted from supporting data from FDA registration documents. The data source table is given in the Appendix in Table 11. This shows one row per outcome per study for each drug and the placebo. Cells shaded in yellow denote raw data extract from the source, in pink exact calculations performed on this data and cells not shaded denote that the number in them needed assumptions to be made in their calculations.

The data in the data source table is used to calculate the outcomes for each drug for each outcome on the scale of measurement given in the value tree. These data are given in the master data summary table in Table 12. It is necessary to measure the benefit and risk outcomes on an absolute scale, as this is the scale that values need to be elicited. However, the studies were performed in different patient populations and at different points in time, so comparing absolute effects between treatment arms directly will be confounding the effect of the drug and the effect of the patient population. It is common practice in biostatistics to assume that the relative treatment or safety effect is robust to changes in patient population. We compare drugs by taking a common placebo population, and finding the relevant absolute outcome given a drug by multiplying outcome in the placebo group with the appropriate relative rate. This is a simple form of network meta-analysis. The placebo group in the natalizumab trial is chosen as the common placebo group in most outcomes. However where there are missing data in natalizumab case, the common placebo group is chosen as the one with available data. For rare adverse events such as PML, herpes viral infections, or seizures, it has been assumed that no event has been observed in the common placebo group, and the unadjusted outcome in the treatment arm is used.

7.3.2 Preference weights

The weighting process was done using a bottom-up (or right-to-left approach given the orientation of our value tree) approach in the hierarchical value tree through decision conferences with the European Multiple-Sclerosis Platform (EMSP).

Firstly, the representatives of the EMSP were asked to take the regulator perspective making the decision in drug licensing on behalf of MS patients, however, the EMSP felt they could not take this perspective and instead gave their own values. The weighting process was organised in two meetings. The first meeting was to give an overview of the project and give instructions on completing the questionnaires. Then the representatives completed the questionnaires individually, between the meetings. The second meeting was a group decision conference. The representatives presented their individual answers to the questionnaires and then came to a group consensus for the weights being elicited.

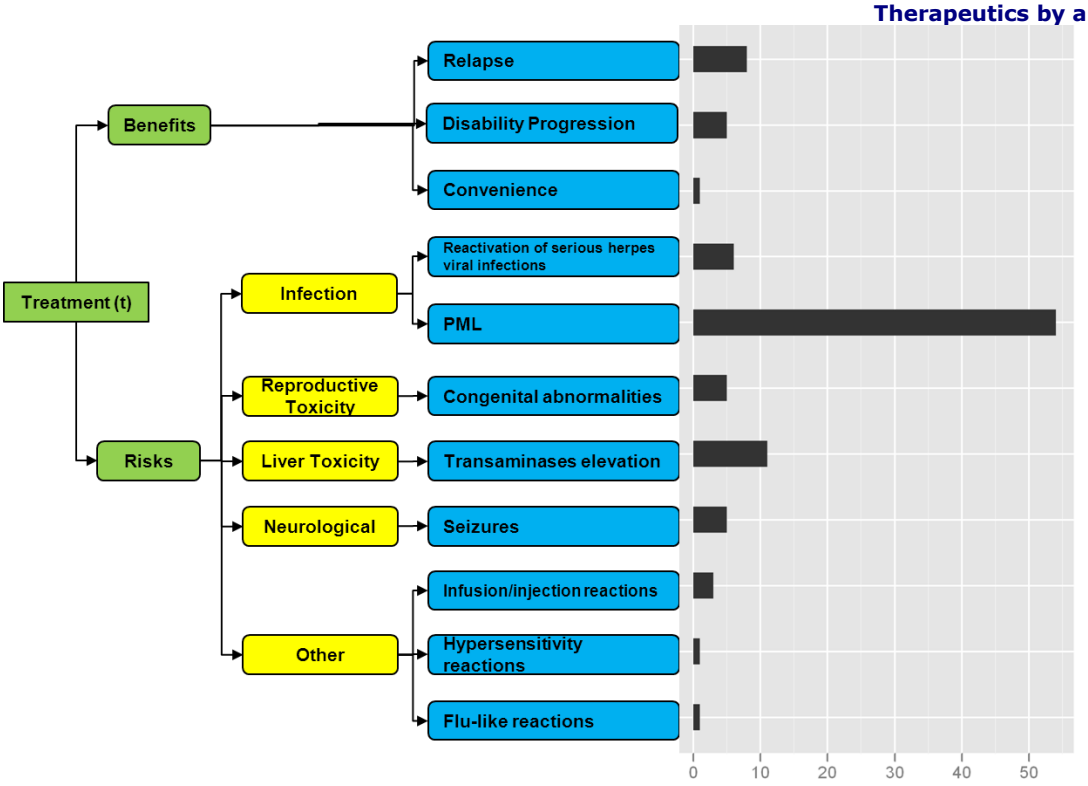
Weights are elicited thought the hierarchy of the value tree. Starting with each outcome measure, the relative ranks of a swing from the best to the worst of each outcome within each category are elected. Start from where all outcome scores are at the worst score, and choose the outcome you would most want to move to the best score. Then rank the other outcomes in a similar way. This is a “thought stepping stone” for then putting preference weights on these outcomes. For this, the top ranked outcomes is given a weight of 100, and place the other outcomes on the scale to reflect their relative importance.

The same approach is used to elicit weights between outcome categories, separately within benefits and risks. The top-ranked outcome from each category is used as a representative of its category. This top-ranked outcome in each category is ranked and weighted in the same way as before. Finally the benefits and risks are weighted using the same approach.

Cross checks were performed, where the weights were converted into the number of patients that would have to have a pair of event avoided to be in equipoise. For example “Imagine a clinical trial of 1000 patients with 1 patient developing PML in the treatment arm. How many patients would need to have an EDSS progression prevented for you to be indifferent about the benefit and harm caused by the treatment?”

The weights elicited at the different levels of the hierarchy are then used to calculate the overall cumulative weights for each outcome. Essentially the weight of an outcome is elicited, and then it is increased in proportion to how much weight it represented lower in the hierarchy. Note that the preference weight for relapses was elicited assuming a range out outcome between 0 and 1. As a range larger than this is needed, it is rescaled to the range 0 to 2, and so the swing weight needs to be doubled to account for this. The weights used for the benefit-risk analysis are shown in 7.3. Note that as the weight for a relapse is for a value function with the measure scale with a range from 0 to 2, then actual weight of a single relapse is half that given in the Figure.

Figure 7.3 Preference weights for each outcome in the value tree



7.3.3 Value function for route of administration

The route of administration outcome is on a categorical scale, and the value of each route and frequency of administration was elicited at the same time as the weights. The elicited value function is shown in Figure 7.4.

Figure 7.4 Value function for route of administration

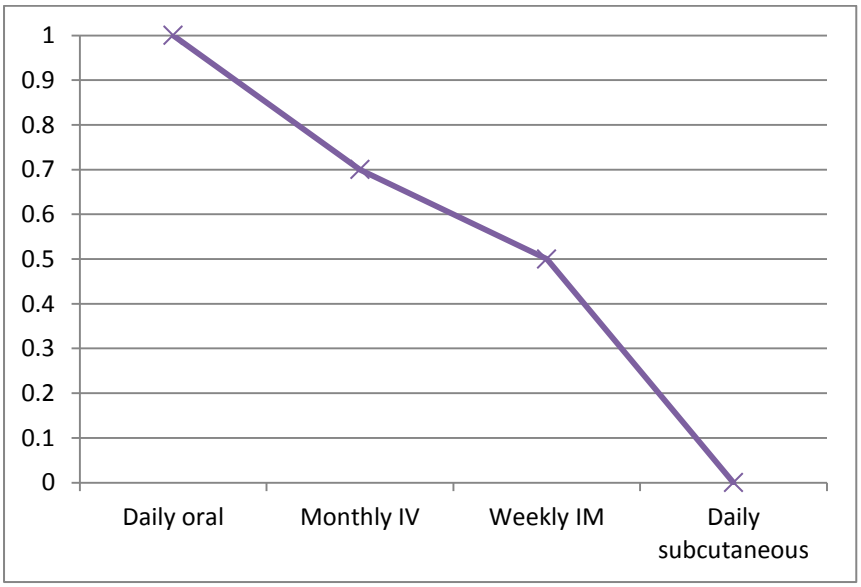


Table 11 Data source table

GROUP	CATEGORY	OUTCOME	STATISTIC	SOURCE	DRUG	ACTIVE			PLACEBO			ACTIVE/PLACEBO		
Benefit						EST	LOW CI		EST	LOW CI	UPP CI	EST	LOW CI	UPP CI
	Relapse		Annualised relapse rate	Polman 2006 and EPAR	Natalizumab	0.23	0.19	0.28	0.73	0.62	0.87	0.32	0.26	0.4
				Jacobs 1996	Beta-interferon	0.67			0.82			0.82	0.56	1.20
				Johnson 1998	Glatiramer acetate	0.65			0.91			0.71	0.47	1.08
	Disability progression		6-month confirmed % progressing after 2 years	EPAR	Natalizumab	11%			23%					
			6-month confirmed hazard	Polman 2006	Natalizumab	0.078			0.174			0.46	0.33	0.64
			6-month confirmed % progressing after 2 years	Jacobs 1996	Beta-interferon	21.9%			34.9%					
			6-month confirmed hazard	Jacobs 1996	Beta-interferon	0.165			0.286			0.58	0.32	1.03
			3-month confirmed % progressing after 2 years	Johnson 1995	Glatiramer acetate	21.6%			24.6%					
			Ratio between 3-month and 6-month	Kappos 2010	Glatiramer acetate	0.71			0.79					
			6-month confirmed % progressing after 2 years		Glatiramer acetate	15.3%			19.4%					
			6-month confirmed hazard	Johnson 1995	Glatiramer acetate	0.111			0.144			0.77	0.41	1.46
		Convenience	Route and frequency of administration	Polman 2006	Natalizumab	iv qm hosp								

				Jacobs 1996	Beta-interferon	im qw						
				Johnson 1995	Glatiramer acetate	sc od						
Risk						ACTIVE			PLACEBO			ACTIVE/PLACEBO
						n	N	%	n	N	%	
	Infection	Reactivation of serious herpes viral infections		Polman 2006	Natalizumab	0	627	0	0	312	0%	1.00
				Jacobs 1996	Beta-interferon	0	158	0	0	143	0%	1
				Johnson 1995	Glatiramer acetate	0	125	0	0	126	0%	1
		PML		Kappos 2011	Natalizumab	5	3722	0.15 10%	0	3722	0%	
				Jacobs 1996	Beta-interferon	0	158	0%	0	143	0%	
				Johnson 1995	Glatiramer acetate	0	125	0%	0	126	0%	
	Reproduction toxicity	Congential abnormalities		Polman 2006	Natalizumab	0	627	0	0	312	0	
				Jacobs 1996	Beta-interferon	0	158	0	0	143	0	
				Johnson 1995	Glatiramer acetate	0	125	0	0	126	0	

	Liver Toxicity	Transaminases elevation	ALT >5x ULN	Polman 2006	Natalizumab	31	627	5%	12	312	4%	1.25
				Jacobs 1996	Beta-interferon	0	158	0%	0	143	0%	1
				Johnson 1995	Glatiramer acetate	0	125	0%	0	126	0%	1
	Neurological	Seizures		Polman 2006	Natalizumab		627	0.50 %		312	0.50 %	
				FDA	Beta-interferon		158	3%		143	0%	
				Johnson 1995	Glatiramer acetate		125	0%		126	0%	
	Others	Infusion reactions/injection reactions		Polman 2006	Natalizumab	148	627	24%	55	312	18%	1.34
				FDA	Beta-interferon	20	158	13%	18	143	13%	1.00
				Johnson 1995	Glatiramer acetate		125	90%		126	59%	1.53
		Hypersensitivity Reactions		Polman 2006	Natalizumab	25	627	4%	0	312	0%	
				Jacobs 1996	Beta-interferon		158	0		143	0	1
				FDA	Glatiramer acetate	17	563	3%	11	564	2%	1.50
		Flu-like reactions		Polman 2006	Natalizumab		627	0%		312	0%	1

Jacobs 1996	Beta-interferon	96	158	61%	57	143	40%	1.52
Johnson 1995	Glatiramer acetate		125	0		126	0%	1

Table 12 Master data summary table

GROUP	CATEGORY	OUTCOME	MEASURE	DRUG	COMMON PLACEBO		RELATIVE VALUE		OUTCOME ON VALUE SCALE
					DRUG	EST	DESCRIPTION	EST	EST
Benefit	Relapse	Relapse	2 year relapse rate	Placebo	Natalizumab PBO	0.73[1]	rate ratio	1	1.46
				Natalizumab		0.73		0.32	0.47
				Beta-interferon		0.73		0.82	1.19
				Glatiramer acetate		0.73		0.71	1.04
	Disability progression	Disability progression	6-month confirmed % progressing after 2 years	Placebo	Natalizumab PBO	0.17	hazard ratio	1.00	0.23
			6-month confirmed % progressing after 2 years	Natalizumab		0.17		0.46	0.11
			6-month confirmed % progressing after 2 years	Beta-interferon		0.17		0.58	0.14
			6-month confirmed % progressing after 2 years	Glatiramer acetate		0.17		0.77	0.18
	Convenience	Convenience	Route and frequency of adminsitration	Placebo					oral od
				Natalizumab					iv qm hosp

				Beta-interferon Glatiramer acetate				im qw sc od
Risk	Infection	Reactivation of serious herpes viral infections		Placebo Natalizumab Beta-interferon Glatiramer acetate	Natalizumab PBO	0% 0% 0% 0%	event ratio	1 1.00 1.00 1.00	0.0% 0.0% 0.0% 0.0%
		PML		Placebo Natalizumab Beta-interferon Glatiramer acetate				0.0% 0.151% 0.000% 0.000%
	Reproduction toxicity	Congenital abnormalities		Placebo Natalizumab Beta-interferon Glatiramer acetate				0.0% 0.0% 0.0% 0.0%
	Liver Toxicity	Transaminases elevation	ALT >5x ULN	Placebo Natalizumab Beta-interferon	Natalizumab PBO	4% 4% 4%	event ratio	1 1.25 1.00	4.0% 5.0% 4.0%

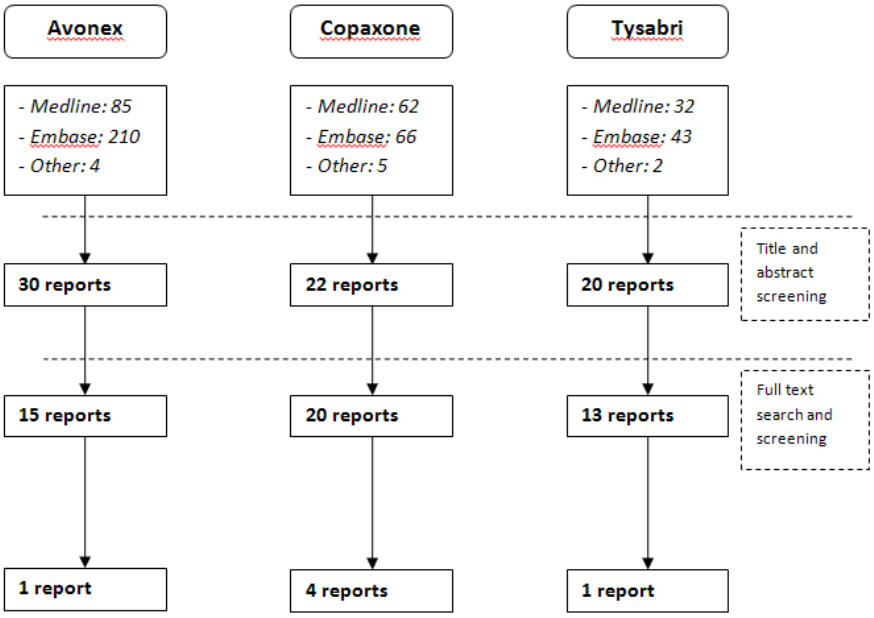
				Glatiramer acetate		4%	1.00	4.0%
	Neurological	Seizures		Placebo				0%
				Natalizumab				0%
				Beta-interferon				3.0%
				Glatiramer acetate				0.0%
	Others	Infusion reactions/injection reactions		Placebo	Natalizumab PBO	18%	event ratio 1	0.0%
				Natalizumab		18%	1.34	23.6%
				Beta-interferon		18%	1.00	17.6%
				Glatiramer acetate		18%	1.53	26.9%
		Hypersensitivity Reactions		Placebo	Natalizumab PBO	0%	event ratio 1.00	0.0%
				Natalizumab		0%	0.00	0.0%
				Beta-interferon		0%	1.00	0.0%
				Glatiramer acetate		0%	1.50	0.0%
		Flu-like reactions		Placebo	Beta-interferon PBO	40%	event ratio 1.00	39.9%
				Natalizumab		40%	1	39.9%
				Beta-interferon		40%	1.52	60.8%
				Glatiramer acetate		40%	1	39.9%

1) One year rate. Value tree outcome is the two year rate so this is doubled

Part 7.4 Phase I of the visualisation methodology

7.4.1 Flow chart 1

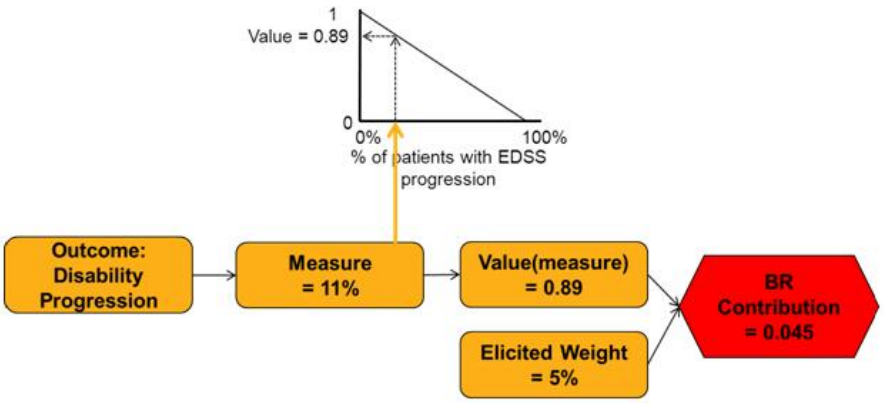
Literature searching results



Name/rubric:	Flow chart showing systematic review literature screening
Created in:	Microsoft Word (drawing facility)
Intended audience:	Statisticians and regulators. Not for patients.
Message:	To visualise the flow of a literature search. Figure shows the sources and process of extracting evidence data. It also shows the amount of relevant data available in terms of number of articles/reports.
Knowledge required:	Some knowledge on systematic review process and quality of database sources.
Unintentional message:	N/A
Message not communicated:	N/A
Proposed improvement:	Provide number of excluded articles and brief reasons for exclusion at each screening stage
Comment	The visual is not taken into Phase II of visual methodology work since it is too general and too simple to benefit from interactive visualisation.

7.4.2 Flow chart 2

Benefit-risk calculations



Name/rubric:	Flow chart
Created in:	MS Powerpoint
Intended audience:	Statisticians, Regulators, not for Physicians or Patients
Message:	To visualise the MCDA concept of scoring. The figure shows benefit risk contribution of one isolated parameter (Disability); relationship between value and parameter, modeled as a linear function.
Knowledge required:	In depth knowledge of MCDA and value functions. Knowledge of parameter/endpoint on the horizontal axis and the plausible values.
Unintentional message:	Unclear. Also, there is no inclusion of sensitivity or uncertainty aspects of the value function.
Message not communicated:	It is also beneficial to focus further on the likely range of the values on the horizontal axis?
Proposed improvement:	It might help some users understand the concept better if similar figures are produced for all key parameters. The figure would benefit from annotations explaining more clearly why a linear function was chosen. The terminologies used are unfamiliar to many users, therefore would require more explanations e.g. the difference between value and weight, and why MCDA requires them.
Comment	The visual is not taken into Phase II of visual methodology work since it is too general and varies from one approach to another. Explaining concepts of benefit-risk approach through visualisation is also beyond the scope of this case study.

7.4.3 Table 1

Descriptions of key data sources

Drug	Author	Year	Exposure	Comparator	Details
Tysabri	Polman CH	2006	28 Months	Placebo	RRMS. Nov 2001: 942 patients - multiple centres and countries. 856 patients (91%) completed the 120-week trial
Avonex	Jacobs LD	1996	24 Months	Placebo	RR MS. Early 1993: 301 patients - 4 clinical centers US.
Copaxone	Johnson KP	1995	24 months	Placebo	RRMS. October 1991-May 1992: 251 randomized. 215 completed 2-year trial. US population.

Name/rubric:	Simple descriptive table on included study articles
Created in:	Microsoft Word (table facility)
Intended audience:	Statisticians, regulators, and physicians. Not for patients
Message:	To lay out descriptive (or quantitative) information in a grid structure. The figure describes the study characteristics; showing larger, more recent, and longer trial was carried out for natalizumab compared to Beta-interferon and Glatiramer acetate. All active drugs used placebo as comparator.
Knowledge required:	Low statistical and low medical knowledge
Unintentional message:	The table implicitly assumes that the quality of evidence for all studies is similar and for comparison to be made, and that the populations are also comparable. This may not always be the case.
Message not communicated:	Characteristics of patients who took part in the trials are not highlighted, implicitly assuming they are comparable.
Proposed improvement:	Display demographics information for patients who took part in the trials and highlight similarities or dissimilarities.
Comment	The visual is not taken into Phase II of visual methodology work since it is too general and too simple to benefit from interactive visualisation.

7.4.4 Table 2

Master data summary table (only partly-shown here)

GROUP	CATEGORY	OUTCOME	MEASURE	DRUG	COMMON PLACEBO DRUG	EST	RELATIVE VALUE DESCRIPTION	EST	OUTCOME ON VALUE SCALE EST
Benefit	Relapse	Relapse	2 year relapse rate	Placebo	Tysabri PBO	0.73[1]	rate ratio	1	1.46
				Tysabri		0.73		0.32	0.47
				Avonex		0.73		0.82	1.19
				Copaxone		0.73		0.71	1.04
	Disability progression	Disability progression	6-month confirmed % progressing after 2 years	Placebo	Tysabri PBO	0.17	hazard ratio	1.00	0.23
			6-month confirmed % progressing after 2 years	Tysabri		0.17		0.46	0.11
			6-month confirmed %						

Name/rubric:	Table as used in BRAT framework (only a part of full table is shown)
Created in:	Microsoft Word (table facility)
Intended audience:	Statistician, regulators and physicians. Not for patients.
Message:	To lay out descriptive (or quantitative) information in a grid structure The figure shows a summary of the master data consisting of benefits and risks criteria used, the way they are measured, and their magnitudes from all studies for all drugs being compared.
Knowledge required:	Some familiarity with the units of measurements and terminologies used – percentage (%), rate ratio, hazard ratio, value scale. Some medical knowledge on the outcomes and epidemiology of the drug-disease.
Unintentional message:	The table header labelling is confusing, particularly in the sixth column when it is labelled with "Common Placebo" but listed natalizumab in subsequent rows in that column. The colour-coding and variation are meaningless since they only represent hierarchy of criteria which does not add value.
Message not communicated:	NA. Note: This is part of a larger table with risks also defined.
Proposed improvement:	Colour-coding should be done by row instead of column to be meaningful. Colour-coding by rows imply grouping of benefits and risks criteria therefore making the measurements more easily interpretable.
Comment	The visual is not taken into Phase II of visual methodology work since it is too general and too simple to benefit from interactive visualisation.

7.4.5 Table 3

natalizumab versus Placebo (Comparator) at time of CHMP re-evaluation

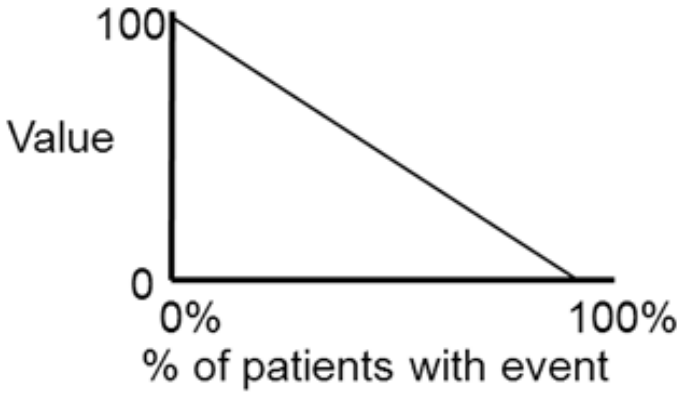
		Outcome	Tysabri Risk / 1000 pts	Comparator Risk / 1000 pts	Risk Difference (95% CI)/ 1000 pts	
Benefits	Convenience Benefits	Convenience (weight 0.6%)	-	-	-	(-, -)
	Medical Benefits	Relapse (weight 3.9%)	280	540	-260	(-326, -195)
		Disability Progression (weight 5.6%)	110	230	-120	(-, -)
Risks	Infection	Reactivation of serious herpes viral infections (weight 6.7%)	80	70	10	(-26, 45)
		PML (weight 55.9%)	2	0	2	(-, -)
	Liver Toxicity	Transaminases elevation (weight 11.2%)	50	40	10	(-16, 38)
	Reproductive Toxicity	Congenital abnormalities (weight 5.6%)	-	-	-	(-, -)
	Neurological Disorders	Seizures (weight 5.6%)	0	0	0	(-, -)
	Other	Infusion/Injection reactions (weight 2.8%)	236	180	56	(6, 114)
		Hypersensitivity reactions (weight 1.1%)	90	40	50	(20, 82)
		Flu-like reactions (weight 1.1%)	399	400	-1	(-114, 114)

Higher for Tysabri
Higher for Comparator

Name/rubric:	Key benefit risk summary table as used in BRAT framework
Created in:	BRAT Excel Tool table with edit in Microsoft PowerPoint
Intended audience:	Physicians and regulators. Not for patients
Message:	To lay out descriptive (or quantitative) information in a grid structure. The figure shows quantitative key benefits and risk with comparison to alternative treatments. The absolute numbers on the table are supportive to allow better judgments.
Knowledge required:	Low statistical knowledge and some medical knowledge of the outcomes.
Unintentional message:	Risks have higher weights than benefits from the number of criteria used in the model. The legend is misleading when the colours in the last column are coded as 'Higher for natalizumab' instead of coding in terms of 'Favors natalizumab', and likewise for Comparator.
Message not communicated:	The criteria weights are displayed in the table but the role of the weights is not clear. It is unclear whether the results presented have taken into account weights or whether the weights should be considered separately. This in turns could make deciding on the benefit risk balance much more critical.
Proposed improvement:	Comparison to other comparators than placebo may also help in the decision-making. A table can be more complex than a graph. The use of blue and yellow scheme should be replaced with the traditional green and red scheme. If weights are to be presented, the ordering should reflect the weights. Weights should be removed or incorporated into the incidences and other calculations since presenting individual weights is not self-explanatory. Labelling of placebo should be explicit as "Placebo" instead of "Comparator". There should also be a text note to emphasise that data come from clinical trials.
Comment	The visual is not taken into Phase II of visual methodology work since it is too general and too simple to benefit from interactive visualisation.

7.4.6 Line graph 1

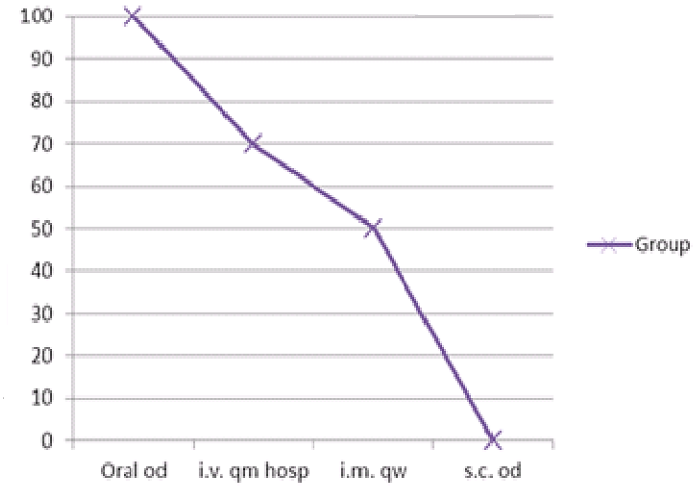
Value function used for proportions



Name/rubric:	Line graph of preference values against a binary event (proportions of patients with an event)
Created in:	Microsoft Word Draw facility. Can also be created in most packages
Intended audience:	Statisticians, Regulators, not for Physicians or Patients
Message:	To show the relationship between preference value and data for one isolated parameter, under linear map assumption of the value function. The representation is unclear, especially when compared to a similar figure in 7.4.2 which includes flow chart for more context and information.
Knowledge required:	Needs in depth knowledge of MCDA and value functions. Knowledge of the parameter/endpoint on horizontal axis and its plausible values are also required.
Unintentional message:	There is no inclusion of sensitivity or uncertainty aspects which are likely to be associated with individual's preference. The aspect ratio of less than 1 (longer vertical axis than horizontal axis), may be perceived as increase in 1% of patients with event translates to less than 1 unit increase in preference value when they actually equal to the same amount.
Message not communicated:	Value function for other benefits and risks criteria are not shown but could easily be produced. There is no explanation whether the value function refers to active treatments or placebo, or is the same for all treatments. It is beneficial to focus further on the likely range of values on the horizontal axis (data values) and the justifications for range of values choice.
Proposed improvement:	The value function should be provided for all key parameters. A text explanation to justify the use of a linear value function is needed for transparency.
Comment	The visual is not taken into Phase II of visual methodology work since it is too simple to benefit from interactive visualisation

7.4.7 Line graph 2

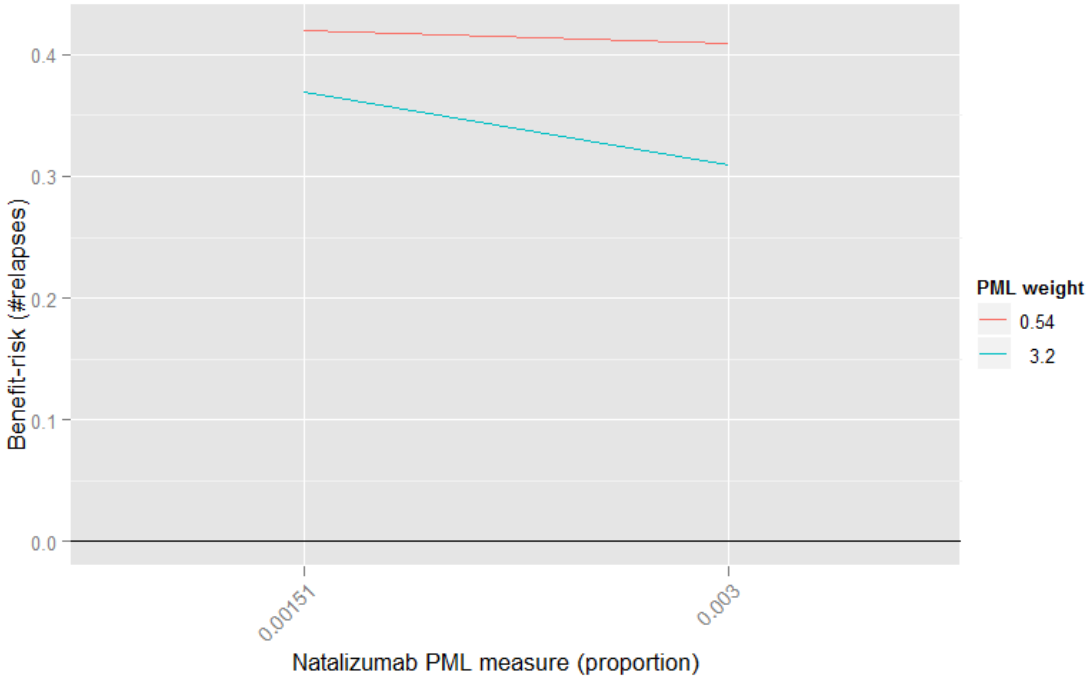
Value function used for convenience (administration route)



Name/rubric:	Value function for convenience (administration route)
Created in:	Micorosoft Excel. Can also be created in most software
Intended audience:	Statisticians, Regulators, not for Physicians or Patients
Message:	Preference values of a categorical variable (administration route), ordered semi-arbitrarily from high preference to low preference.
Knowledge required:	The terminology used, in this case route of administration and the abbreviations, need to be understood to make sense of the graph.
Unintentional message:	Erroneously displays a categorical variable as if it were continuous and for the untrained eye seems to assume a natural ordering, which might in reality be considered post-hoc.
Message not communicated:	The role of the legend “Group” is not communicated and is misleading since there is no grouping to be seen on the graph. Also, see below and above.
Proposed improvement:	Use bar chart or dot plot instead, making clear that the variables are categorical not continuous. At the very least, the connecting line needs to be removed.
Comment	The visual is taken into Phase II visual methodology work as an example of the likely improvement for the representation of a discrete/categorical variable.

7.4.8 Line graph 3

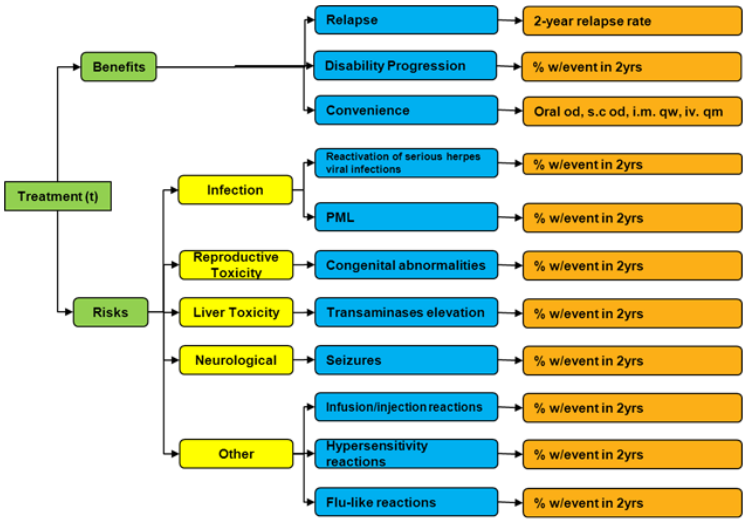
Two-way sensitivity analysis plot



Name/rubric:	Two-way sensitivity analysis plot.
Created in:	R
Intended audience:	Statisticians and regulators.
Message:	To show how the changes in both the number of patients developing PML and the weight associated with PML affect the benefit-risk score. It shows the sensitivity of benefit-risk balance, for values of a variable (here % of patients with PML) for different weights of that parameter.
Knowledge required:	Need to understanding of concept of weights, as well as know the probabilities of PML events. Users also need to know that negative BR values represent 'poorer' outcomes.
Unintentional message:	Low "PML weight" might be interpreted as lower preference to experiencing PML.
Message not communicated:	The choice of particular weight values is unclear from the graph. There is also no mention that the lines only represent benefit-risk scores against plausible proportions of patients who may experience PML with treatment.
Proposed improvement:	The graph could be somewhat confusing to someone with a lack of background knowledge on the problem at hand; perhaps more explanation in the way of a title or annotation is required. Explanation of weights choice is also required. More distinct colours should be used to discriminate the effects of different weights.
Comment	The visual is not taken into Phase II of visual methodology work since it is too simple to benefit from interactive visualisation

7.4.9 Value tree 1

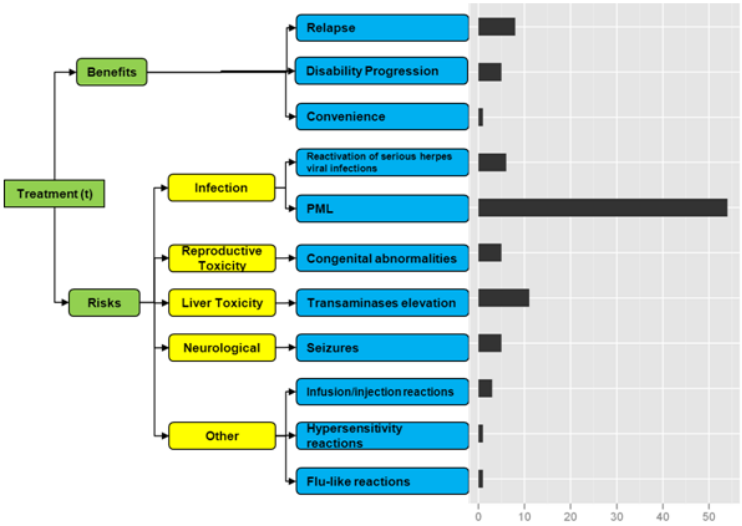
Value tree from BRAT framework



Name/rubric:	Value tree
Created in:	BRAT Lotus notes program plus PowerPoint, but can also be created in PowerPoint alone
Intended audience:	Regulator and Physician. Also for patients if background is provided on the medical terms in lay language.
Message:	To display qualitative listing of available key benefits and risks criteria for the decision model and the description of their measurements in a hierarchical way. Colours are used to indicate level of criteria.
Knowledge required:	Needs Medical knowledge (terminology of the medical terms; for judging whether selected benefits and risks make sense), no statistical knowledge needed. The tree could be amended for the patients as well using lay terminology to be easily readable.
Unintentional message:	As usually benefits are predetermined, and on the risk side, potential risks are also included so there is a tendency to show more risks than benefits. For the inexperienced reader this could be perceived as “more” risks if no further qualification of the risks is provided. This could be done by more explicit description of certainties, numbers are not weighted but weights could be added alongside. As presented, each criterion may be seen as being equally weighted.
Message not communicated:	The context in which the value tree is created is needed (underlying database, indication, what decision). There is no quantitative data and no information on the treatment alternatives.
Proposed improvement:	Some background information, such as definitions for the indications should be included via text annotation. Explanation of whether the order of criteria reflects weighting would be useful. The naming of criteria must intuitive reflect their role as a benefit or a risk of treatment, for example write “relapse” as “reduction in relapse rate” to reflect it is a benefit of treatment. Different colours should be used to distinguish benefit and risks. Risks should be characterised in terms of certainty, and lay terms to be added to make the value tree more useful for patients.
Comment	The visual is taken into Phase II of visual methodology work to illustrate how text annotations and colour choices may help improve the understanding of a value tree in an interactive visualisation

7.4.10 Value tree 2

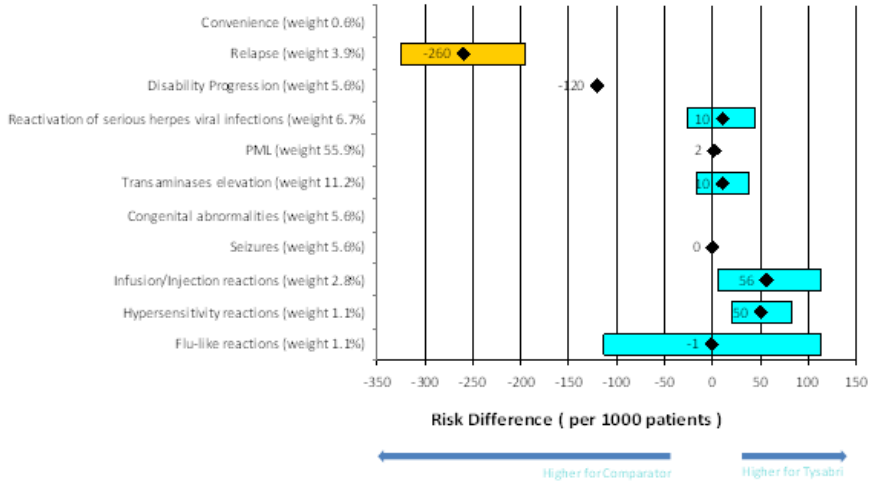
Value tree from BRAT framework with added preference weights



Name/rubric:	Value tree with preference weights
Created in:	BRAT Lotus notes program plus power point
Intended audience:	Regulators , Physicians, not patients
Message:	To display qualitative listing of available key benefits and risks criteria for the decision model and the description of their measurements in a hierarchical way. Patient preferences on each criterion are also displayed. Colours are used to indicate level of criteria.
Knowledge required:	Needs Medical knowledge (terminology of the medical terms; for judging whether selected benefits and risks make sense), no statistical knowledge needed. Understanding of the meaning and role of weights on the value tree is required.
Unintentional message:	Risks are weighted very high compared to benefits. Since the probability of event is missing, individuals may quickly judge that risk clearly outweighs benefit.
Message not communicated:	The way parameters were measured in the trial (ALT Units etc.) is unclear (but included in value tree in Section 7.4.9. It is also unclear which treatments are being compared, and whose preference weights are being represented.
Proposed improvement:	The axis for preference weights should be labelled and the scale is to be made explicit. Whose preference weights are being represented should be made clear through the axis label. The criteria could be ordered by magnitude of weights within the hierarchy. The graph should be accompanied by the frequency of events to allow for appropriate judgment to be made.
Comment	The visual is not taken into Phase II of visual methodology work since it is too similar to value tree 1 in Section 7.4.9, which has been carried forward.

7.4.11 Dot/Forest plot

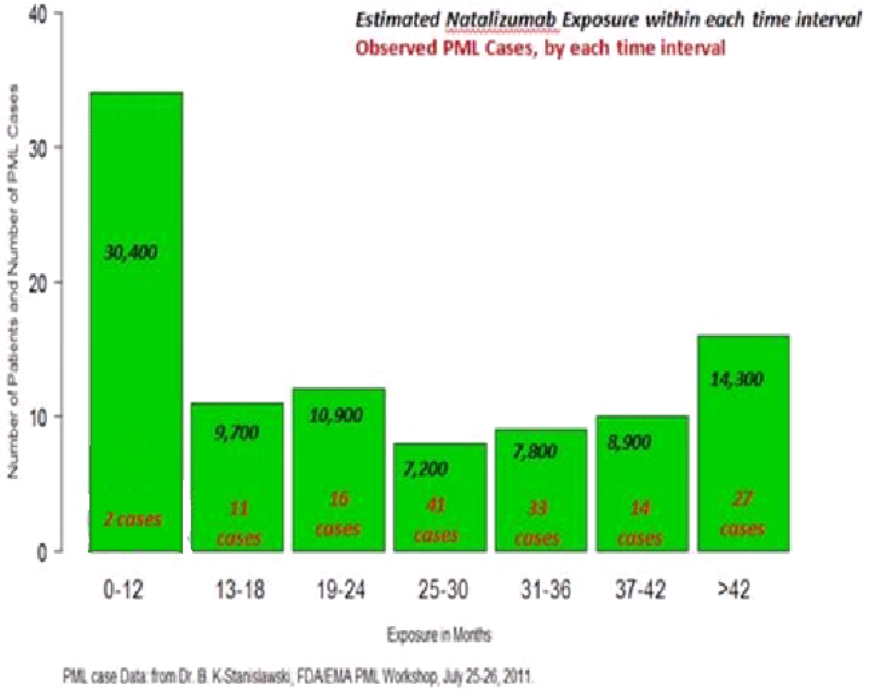
Forest plot from BRAT framework



Name/rubric:	Forest plot on absolute risks
Created in:	BRAT Excel Tool adapted in PowerPoint, but could also be done in PowerPoint alone.
Intended audience:	Statisticians, Physicians, Regulators; not for patients
Message:	It shows the risk difference between natalizumab and placebo at time of CHMP re-evaluation on individual key benefits and risks criteria. The forest plot leads to judgement of positive benefit risk balance. Absolute numbers as presented are more supportive to allow judgements when compared to relative values. Statistical uncertainty of the risk difference is also given. The forest plot is a simple visual graph, and is easier to comprehend when compared to a table with the same information.
Knowledge required:	Needs medical knowledge, low statistical knowledge
Unintentional message:	Risk of very small but important events may be undervalued
Message not communicated:	The weights appeared with criteria name are confusing since it is unclear how they are to be taken into account in this graph. “Convenience” and “Congenital abnormalities” criteria appear on the forest plot but are not quantified. In general, non-categorical values are difficult to capture i.e. 6 min walk test, time aspect not covered as for example in Kaplan Mayer curves
Proposed improvement:	Replace “higher to” by “favours” if the latter can correctly describe the direction, or otherwise describe whether “higher” is desirable or undesirable outcome. Use green and red instead of blue and yellow with clearer text colours (black is likely to be the most visible here). The plot could have dual horizontal axis to represent both continuous and categorical outcomes. A second graph can be added where weights have been incorporated into the measure of differences.
Comment	The visual is taken into Phase II of visual methodology work to demonstrate how a forest plot can be used more efficiently as an interactive graph.

7.4.12 Bar chart 1

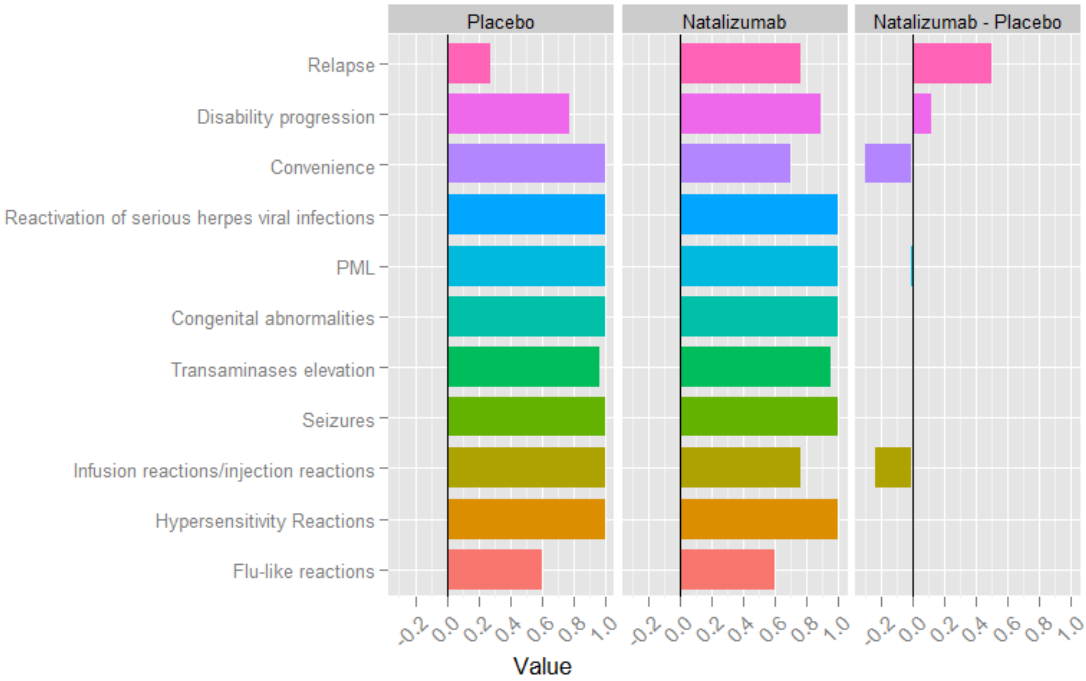
A simple bar chart



Name/rubric:	Simple bar graph
Created in:	Extracted from external source and edited in Microsoft Word
Intended audience:	All
Message:	The bar chart communicates the distribution of PML cases from post-marketing exposure of natalizumab by number of months of exposure. A short exposure (0-12 months) to natalizumab is the most common. The rate of PML cases increases with longer use up to three years.
Knowledge required:	Low statistical and medical knowledge. Users need to understand that only the height of the bar is to be interpreted not the area.
Unintentional message:	The bar value label does not match the values on the vertical axis which would confuse the users. It gives the impression that those in 25-30 months category had the lowest PML rate. The width of the bars may be confused as being of the same range.
Message not communicated:	It is unclear whether "PML cases" is the number of PML incidence or number of patients who experienced PML regardless of number of PML events.
Proposed improvement:	Colours of bars and labels should be chosen more carefully to increase contrast so that users would be able to read them easily. Stacked bar graphs (PML + non PML) of the percentage may be more suitable to represent the information in the current graph.
Comment	The visual is not taken into Phase II of visual methodology work since it is too simple to benefit from interactive visualisation

7.4.13 Bar chart 2

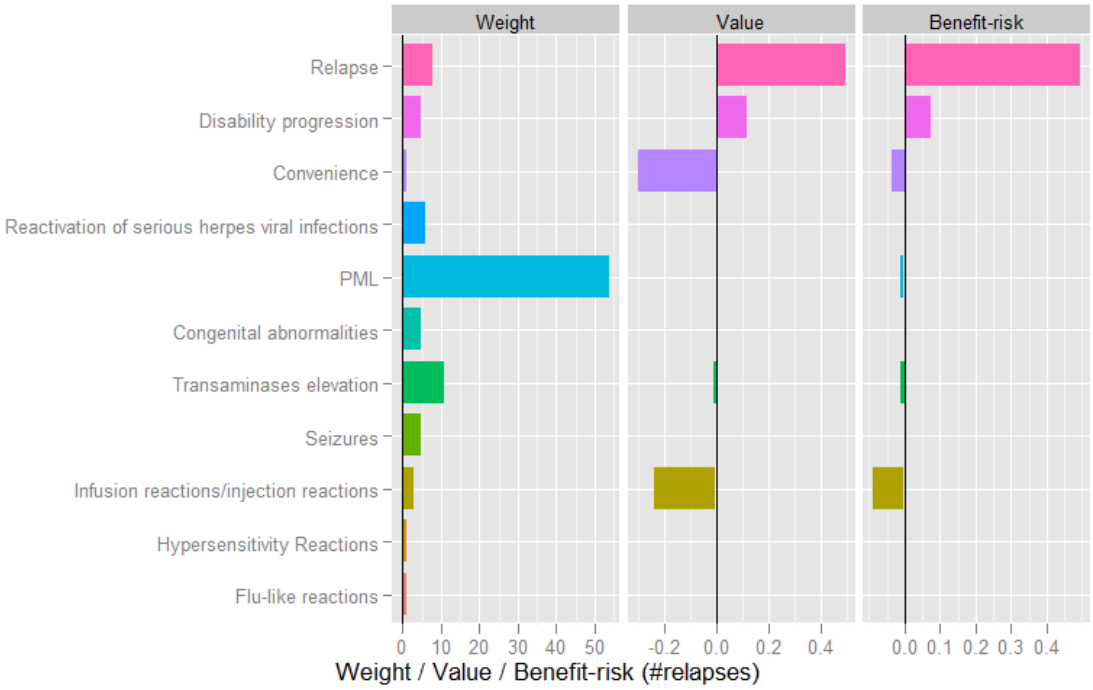
Aligned bar chart of utility values by treatment and their difference



Name/rubric:	Aligned bar chart with difference display
Created in:	R (ggplot2 package)
Intended audience:	Statisticians, Regulators not for Physicians or patients
Message:	The bar chart shows the quantitative key benefits and risks values by criterion and comparative treatments, and the difference between them. The meaning of the values is unclear. The lengths of the bars appear to be the same.
Knowledge required:	Needs in depth knowledge of MCDA and value functions to understand the message properly. But no specific technical knowledge is required to determine on which criteria natalizumab is valued higher or lower than placebo from the difference display (right-most column)
Unintentional message:	There is a lack of transparency and gives an impression of complexity. There is very minimal benefit risk balance. The importance of PML may be underestimated from the value difference.
Message not communicated:	More explanation on whose values and what do the values mean are needed. Statistical uncertainty is not described
Proposed improvement:	Harmonise colours to give meaningful message. Add sensitivity analysis. The meaning of values and their difference need to be explained to aid interpretation. The horizontal axis on the difference display should be labelled with 'Favors natalizumab' and 'Favors Placebo' appropriately.
Comment	The visual is not taken into Phase II of visual methodology work since the real message is quite confusing and may not be very useful for benefit-risk assessment. Alternative aligned bar chart is described in Section 7.4.14.

7.4.14 Bar chart 3

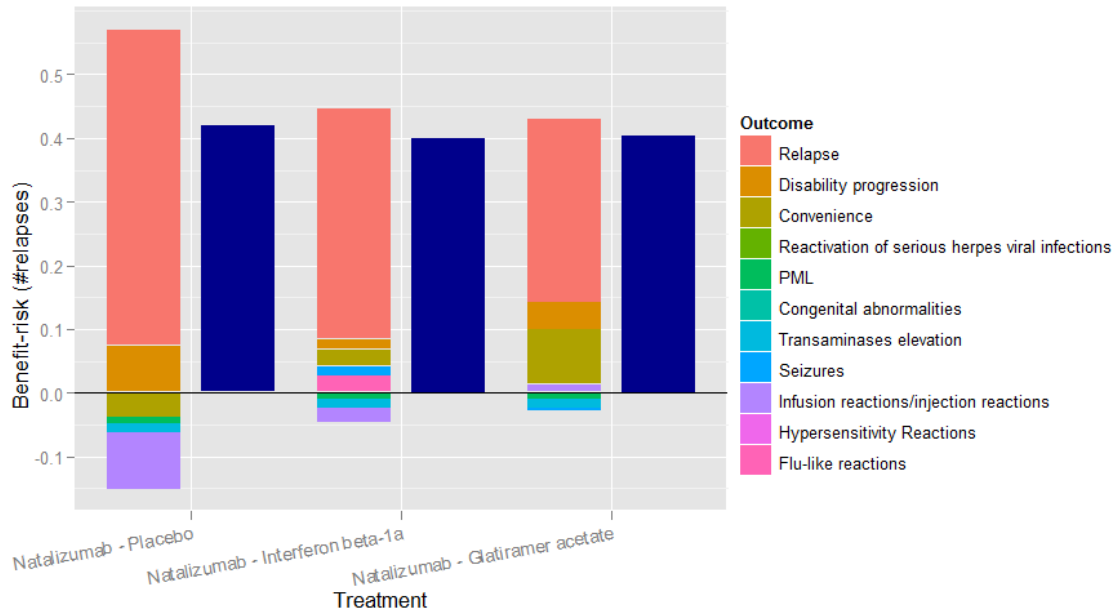
Aligned bar chart of the derivation of the difference in weighted values



Name/rubric:	Bar Chart (of Weights, Values and Benefit Risk Scores)
Created in:	R (ggplot2 package)
Intended audience:	Statistician, Regulators, Physicians, Patients
Message:	The bar chart shows the effect of combining specific weights with difference in values for the quantitative key benefits and risks comparing natalizumab to placebo.
Knowledge required:	Needs in depth knowledge of MCDA and value functions or a lot of trust. But no specific technical knowledge is required to determine on which criteria natalizumab performs better or poorer than placebo from the difference display (right-most column)
Unintentional message:	Rare severe risks may be underestimated.
Message not communicated:	Statistical uncertainty is not communicated making the results seem too certain. Confidence intervals from clinical study data have not been considered; thus results may not reflect the clinical data in whole.
Proposed improvement:	Add sensitivity analysis. Confusion by normalisation – relapse was scaled to 1, so needs better explanation. Colours to be harmonised to provide more meaningful message.
Comment	The visual is taken into Phase II of visual methodology work since it the representation is found to be useful during decision conferencing in Wave 1 case study, but could be improved further to allow better user comprehension.

7.4.15 Bar chart 4

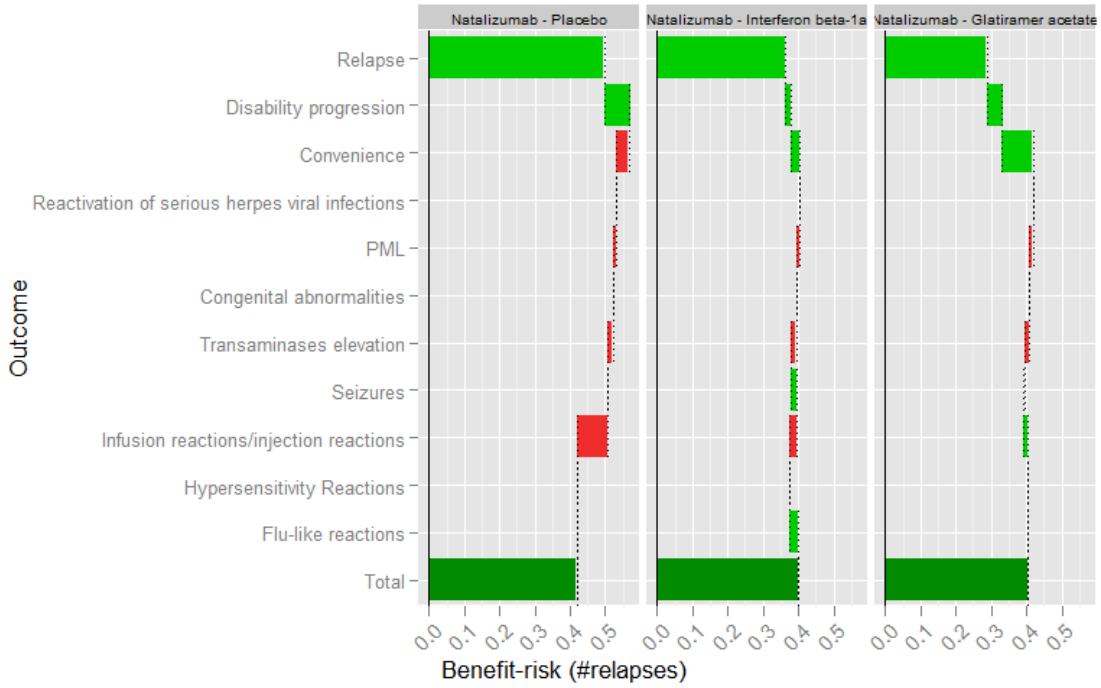
A stacked bar chart of incremental benefit-risk



Name/rubric:	Stacked bar chart (of incremental benefit risk)
Created in:	R (ggplot2 package)
Intended audience:	All
Message:	The bar chart shows the incremental benefit-risk contribution by criterion for natalizumab compared to placebo, Beta-interferon and Glatiramer acetate. It displays the magnitude and direction of contributed criteria to the overall score. It also displays the magnitude and direction of the change in benefit values having discounted risk by criteria and treatment (dark blue bar).
Knowledge required:	No specific technical knowledge is required. Some understanding of the incremental benefit-risk concept i.e. the knowledge that negative values equate to decreased benefit-risk balance, and positive values equate to increased benefit-risk balance.
Unintentional message:	No uncertainty in the benefit-risk balance is presented.
Message not communicated:	Very small scores may not be visible, giving the impression that natalizumab exactly equals comparator on the criterion, which may lead to misinterpretation to users. The dark blue bars are not labelled, so the quantity as to what is presented is unknown.
Proposed improvement:	Dark blue bars to be labelled.
Comment	The visual is not taken into Phase II of visual methodology work since it is too similar to bar chart in Section 7.4.16 which was found to be more useful.

7.4.16 Bar chart 5

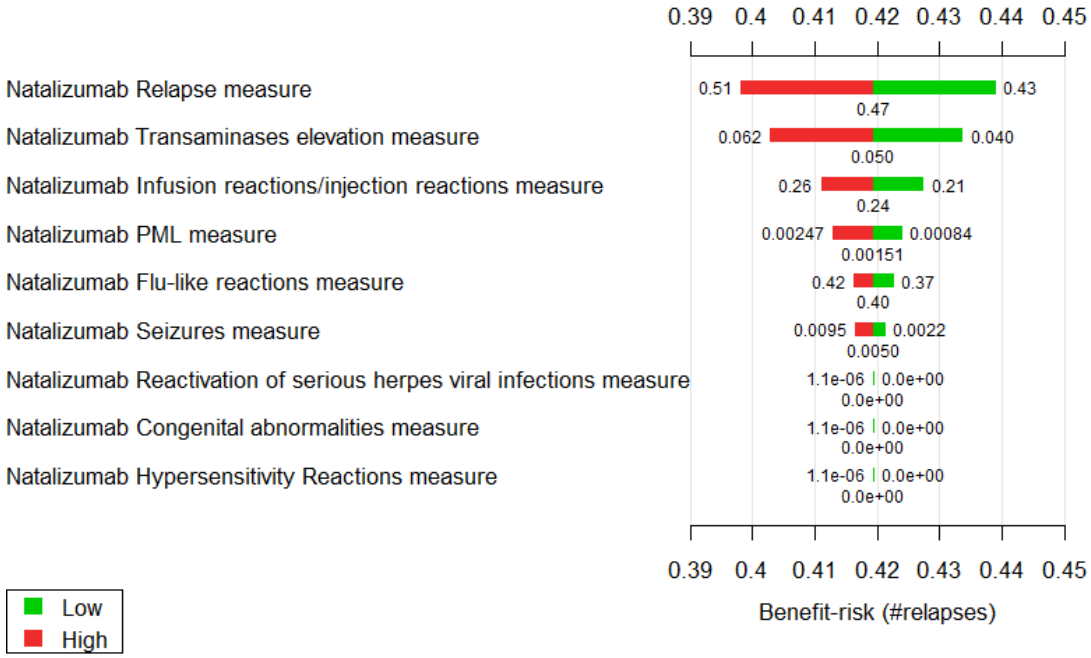
Aligned bar chart of the differences between natalizumab and comparators represented as waterfall plot



Name/rubric:	Waterfall plot (of incremental benefit risk)
Created in:	R (ggplot2 package)
Intended audience:	All
Message:	The waterfall plot shows the incremental benefit-risk of natalizumab compared to placebo, Beta-interferon and Glatiramer acetate. It displays the magnitude and direction of benefit-risk contribution to the overall score by criterion.
Knowledge required:	Some understanding of incremental benefit-risk concept. Some knowledge of waterfall plot e.g. which end of the bars to be read (green = right, red = left). The grids act as reference lines for comparison across panels.
Unintentional message:	Some bars are very narrow – users can't tell whether information is missing or very small values, or the direction of the difference. Inexperience users may find it difficult to extract what the final benefit-risk balance is e.g. green at the bottom may be perceived as positive balance, and red as negative balance which may not always be true.
Message not communicated:	Very small scores may not be visible, giving the impression that natalizumab exactly equals comparator on the criterion, which may not be true since it could be either way and may lead to misinterpretation to users. The overall value is not explicitly communicated.
Proposed improvement:	The horizontal axis should be made wider to enhance readability since the benefit-risk values are represented along this axis. Overall scores should be presented. Ordering of criteria should be more meaningful.
Comment	The visual is taken into Phase II of visual methodology work since it was found to be a useful visualisation tool alternative to stacked bar chart, but require more space to represent the same amount of information.

7.4.17 Bar chart 6

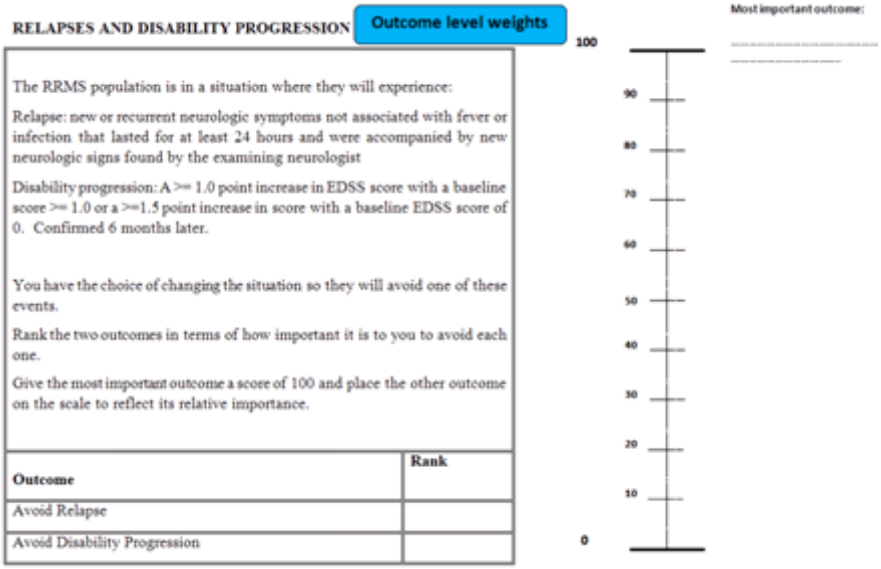
Tornado plot



Name/rubric:	Tornado plot
Created in:	R
Intended audience:	Statisticians and regulators. Not for physicians and patients.
Message:	The tornado plot shows how the changes in the natalizumab outcome measure affect the incremental benefit-risk score. It displays the relative importance of criteria via one-way sensitivity analysis of changing a fixed amount of the measured outcomes.
Knowledge required:	Some knowledge on the use of sensitivity analysis and uncertainty. Some understanding of the incremental benefit-risk concept. Some knowledge on how to extract information from tornado diagrams.
Unintentional message:	The legend of “high-low” is not intuitive and could be misleading.
Message not communicated:	It is unclear which of the criteria are benefits and which are risks. The colour-coding is not intuitive and difficult to interpret.
Proposed improvement:	Horizontal axis should be made wider to accommodate benefit-risk values. To re-label legend items to more intuitive terms. The tornado plot could also be accompanied by text annotations to aid interpretation.
Comment	The visual is taken into Phase II of visual methodology work as a method to visually represent deterministic uncertainty.

7.4.18 Thermometer scale

Thermometer scale



Name/rubric:	Thermometer scale
Created in:	Microsoft Word
Intended audience:	Any stakeholders
Message:	Not applicable. It is used to elicit stakeholder's relative importance on criteria
Knowledge required:	Some medical knowledge of the condition/event, in this case relapse rate and disability progression scoring.
Unintentional message:	Not applicable.
Message not communicated:	Not applicable. However, some stakeholders may want to provide values with uncertainty which could be accommodated.
Proposed improvement:	If the elicitation is not done in a facilitated session, clearer instructions are needed.
Comment	The visual is not taken into Phase II of visual methodology work since Phase II focuses on displaying the results than for eliciting preference values.

Part 7.5 Phase II of the visualisation methodology

7.5.1 Discrete data representation

Audience: All (Regulators, physicians, patients, industry)
Communication: To communicate part-to-whole information on the magnitude of value preference for a categorical criterion (administration)
Type of plots: Bar graph / short line bar
Annotation: Texts should be sans-serif fonts (Arial, Tahoma or Verdana), size 10+ points, colour black, and bold for visibility
Axes: Vertical axis is labelled with fixed scale of 0-100.
Styles: Bar graph to show the size of magnitude but contains redundant information, or alternatively short line to indicate the magnitude (higher data-ink ratio)
Colours: Same colour for all bars/short lines (blue)
Content/storylines: - first image: Bar graph of the value preference - filter: None - animation: Allow user to select bar display or short line display via drop-down box. Tooltips are made available to display the definition of each route of administration when user hover cursor over the bars. - drill-down: Not applicable
Data structure: Variables in the data file are *entity* – the full text name of the discrete data point *code* – text abbreviations of the data point (optional) *value* – numeric value of the data point *defn* – text definition of the data point

Figure 7.2 Bar graph representing discrete data

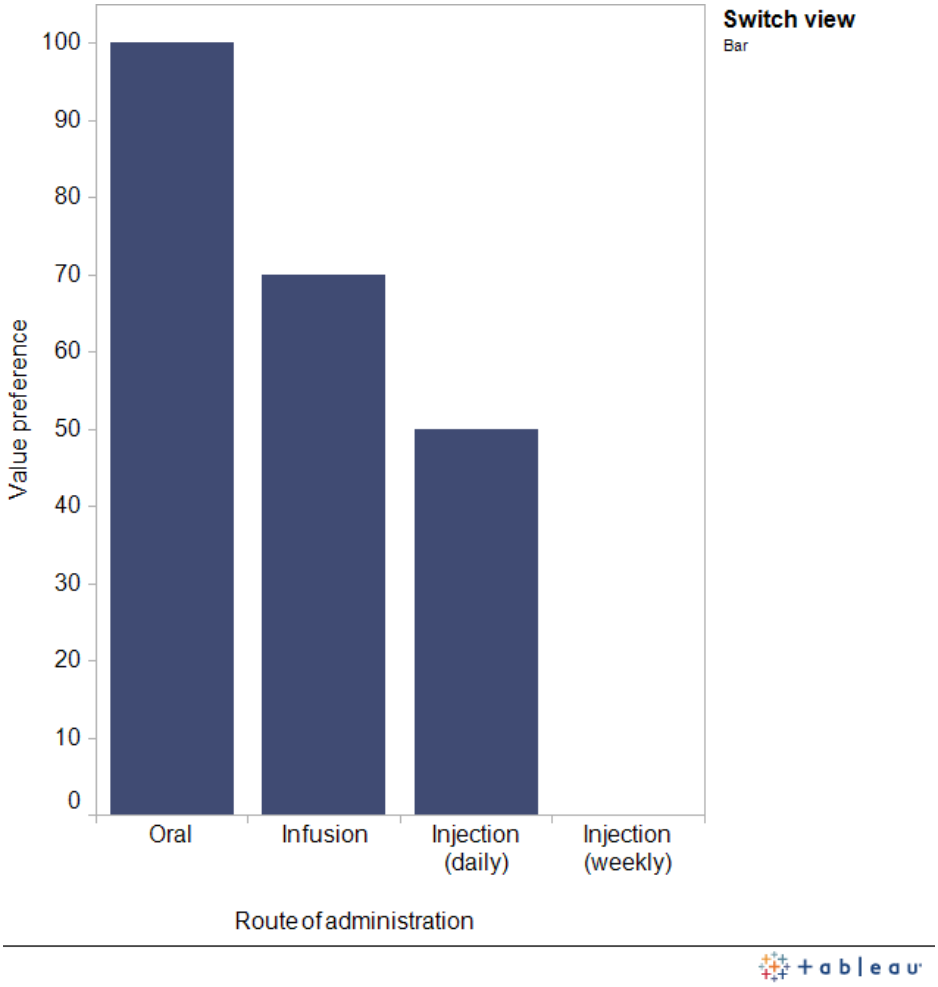
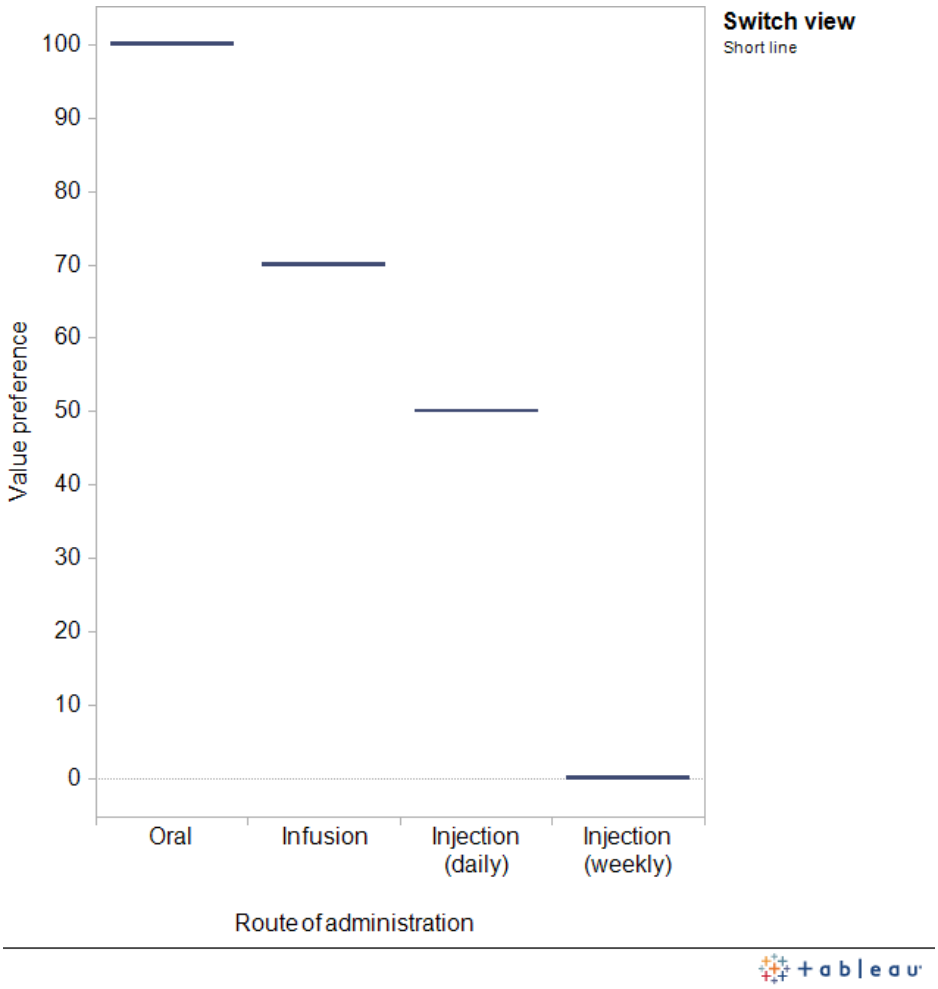


Figure 7.3 Short line graph (dot chart with short line as symbol) representing discrete data



http://public.tableausoftware.com/views/T_Discrete/Discrete

7.5.2 Value tree diagram

Audience: Regulators and physicians, patients if adapted
Communication: To list the criteria for benefits and risks qualitatively by their hierarchy to imply grouping
Type of plots: Tree diagram
Annotation: Texts should be sans-serif fonts (Arial, Tahoma or Verdana), size 10+ points, colour black, and bold for visibility
Axes: Not applicable
Styles: Use oval for decision node and rectangles for criteria. Name the criteria as “reduction in relapse rate”, “slowdown in disability progression” to describe benefits.
Colours: Use green for benefit and red for risk. Administration is given different colour (turquoise) to distinguish from the others. Lower colour saturation by hierarchy to indicate levels.
Content/storylines: - first image: Full tree since it is small and manageable - filter: Filter not applicable - animation: The branches are collapsible when the parent node is clicked. Does not add value here because tree is small. - drill-down: Allow to drill down to the definition and measures. Are displayed when the node is selected. This should only be done at the lowest level criteria. Weights, uncertainties and patient lay language could be added at second level
Data structure: There is no associated data file. The tree diagram is created manually.

Figure 7.4 Tree diagram showing the criteria used in the model and their hierarchy

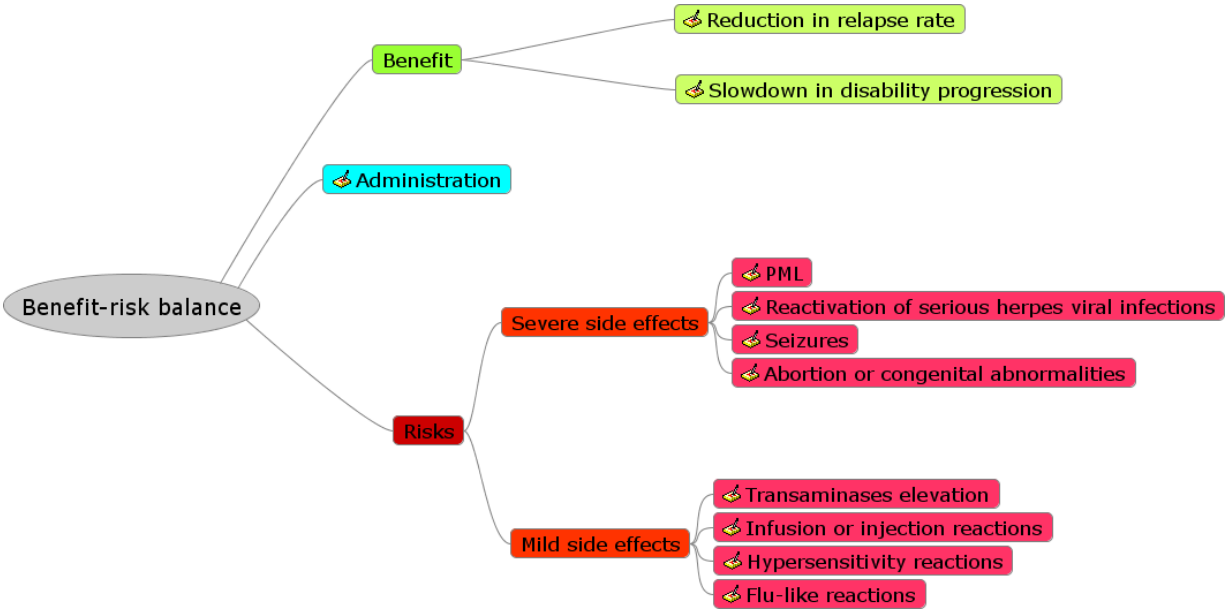


Figure 7.5 Example tooltip defining the criteria (e.g. shown "slowdown in disability progression") to assist user's understanding

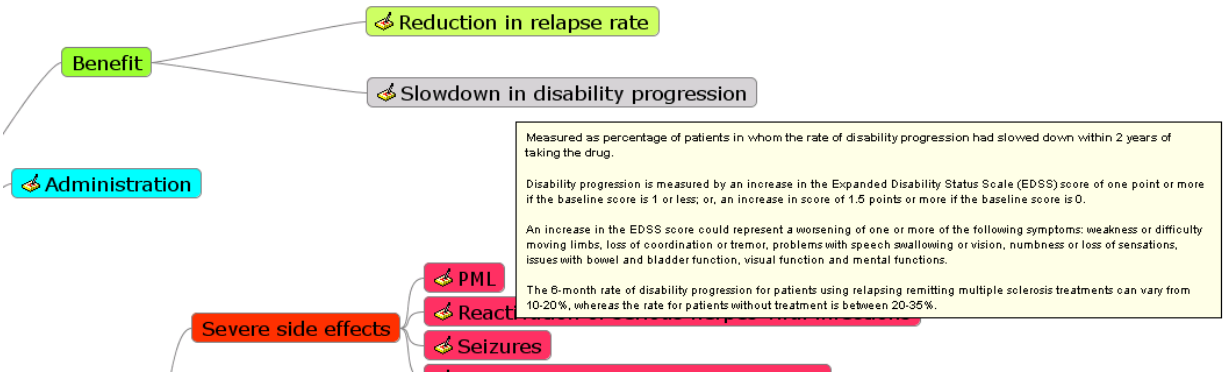


Figure 7.6 Full text descriptions used in the tooltips of definition

Benefit criteria
Administration criteria
Risk criteria

	Criteria	Tooltip's annotations
	Reduction in relapse rates	Measured as 2-year relapse rate A relapse is the appearance of new symptoms, or the return of old symptoms, for a period of 24 hours or more – in the absence of a change in core body temperature or infection. In relapses, symptoms usually come on over a short period of time – over hours or days. They often stay for a number of weeks, usually four to six, though this can vary from very short periods of only a few days to many months. Relapses can vary from mild to severe. At their worst, acute relapses may need hospital treatment, but many relapses are managed at home, with the support of a general practicioner, MS specialist nurse, and/or other care professionals. Symptoms which come and go can sometimes be considered a relapse – they don't always have to be continuous. For example, some people experience a shock-like sensation when they bend their neck. This can be considered a relapse if it occurs every time they bend their neck for at least 24 hours. The relapse rate for patients using relapsing remitting multiple sclerosis treatments can vary from 20-70%, whereas patients without treatment have a relapse rate between 70-85%.
	Slowdown in disability progression	Measured as percentage of patients in whom the rate of disability progression had slowed down within 2 years of taking the drug. Disability progression is measured by an increase in the Expanded Disability Status Scale (EDSS) score of one point or more if the baseline score is 1 or less; or, an increase in score of 1.5 points or more if the baseline score is 0. An increase in the EDSS score could represent a worsening of one or more of the following symptoms: weakness or difficulty moving limbs, loss of coordination or tremor, problems with speech swallowing or vision, numbness or loss of sensations, issues with bowel and bladder function, visual function and mental functions. The 6-month rate of disability progression for patients using relapsing remitting multiple sclerosis treatments can vary from 10-20%, whereas the rate for patients without treatment is between 20-35%.

	Administration	Various different methods can be used to administer medicines. • A capsule to be swallowed once a day, every day. • A once monthly one hour intravenous infusion followed by one hour reaction monitoring at a hospital. • A daily self administered injection that goes just under the skin (subcutaneous). • A weekly self administered injection that goes into the muscle (intramuscular).
	PML	Measured as percentage of patients who experienced PML within 2 years of taking the drug. PML stands for progressive multifocal leukoencephalopathy, a rapidly progressive, often fatal viral infection of the brain. The signs and symptoms of PML include headaches, memory loss, changes in mental status, speech and vision difficulties, loss of strength, limb weakness, seizures, partial paralysis and loss of coordination. The disease leads to coma and then to death. The rate of PML is between 0.0-0.5% across different treatment treatments for relapsing remitting multiple sclerosis. Note that 0.5% is equivalent to 1/200.
	Reactivation of serious herpes viral infections	Measured as percentage of patients who had reactivation of serious herpes viral infections within 2 years of taking the drug. Herpes viruses may become active, e.g. treatment may result in cold sores, shingles, genital herpes, and herpes dermatitis. The rate of reactivation of serious herpes viral infections is between 0-10% across different treatment treatments for relapsing remitting multiple sclerosis.
	Seizures	Measured as percentage of patients who experienced seizures within 2 years of taking the drug. Seizures are caused by uncontrolled electrical activity in the brain. They can range from a wild thrashing bodily movement to a brief loss of awareness. Uncontrolled seizures can cause brain damage, lowered intelligence, and permanent mental and physical impairment. The rate of having seizures is between 0-2% across different treatment treatments for relapsing remitting multiple sclerosis.
	Abortions or congenital abnormalities	Measured as percentage of patients who had abortion or congenital abnormalities to their child. Congenital abnormalities are physical defects present in a baby at birth that can involve many different parts of the body, including the brain, heart, lungs, liver, bones, and intestinal tract. Examples include heart defects, cleft lip and palate, spina bifida, limb defects, and Down syndrome. The rate of having congenital abnormalities is between 0.0-0.5% across different treatment treatments for relapsing remitting multiple sclerosis. Note that 0.5% is equivalent to 1 case out of 200 women of childbearing age receiving the medication.
	Transaminases elevation	Measured as percentage of patients with elevated level of transaminases within 2 years of taking the drug. Transaminases are a type of liver enzyme. An increase in transaminase is unlikely to produce immediate symptoms. However, it can indicate liver damage. The rate of transaminases elevation is between 0-5% across different treatment treatments for relapsing remitting multiple sclerosis

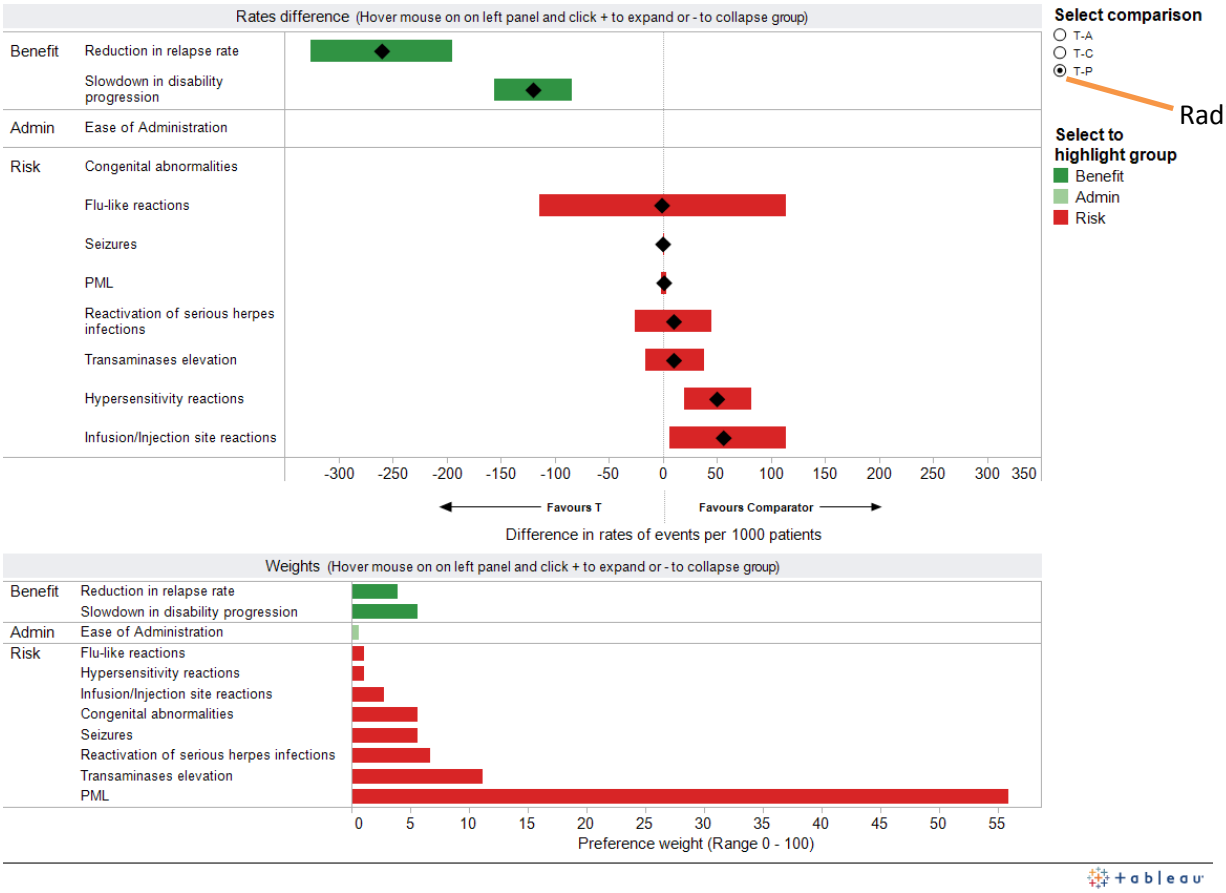
	Infusion or injection site reactions	Measured as percentage of patients who had reactions to infusion or injections sites within 2 years of taking the drug. Infusion or injection site reactions are when the skin reacts to treatment during or after treatment administration. There may be redness, irritation, swelling, and/or pain. The rate of infusion/injection site reactions is between 20-90% across different treatment treatments for relapsing remitting multiple sclerosis.
	Hypersensitivity reactions	Measured as percentage of patients who had hypersensitivity reactions within 2 years of taking the drug. Hypersensitivity reactions are similar to infusion/injection site reactions. However, whereas infusion/injection site reactions affect only the skin, hypersensitivity reactions also affect the rest of the body through an allergic immune response. In rare severe cases, it may result in difficulty breathing and collapse. The rate of hypersensitivity reactions is between 0-10% across different treatment treatments for relapsing remitting multiple sclerosis.
	Flu-like reactions	Measured as percentage of patients who had any flu-like reactions within 2 years of taking the drug. Flu-like reactions are symptoms similar to experiencing flu, e.g. fever, chills, fatigue, nausea and vomiting. The rate of flu-like reactions is between 40-60% across different treatment treatments for relapsing remitting multiple sclerosis.

Benefit criteria
Administration criteria
Risk criteria

7.5.3 Dot or forest plot

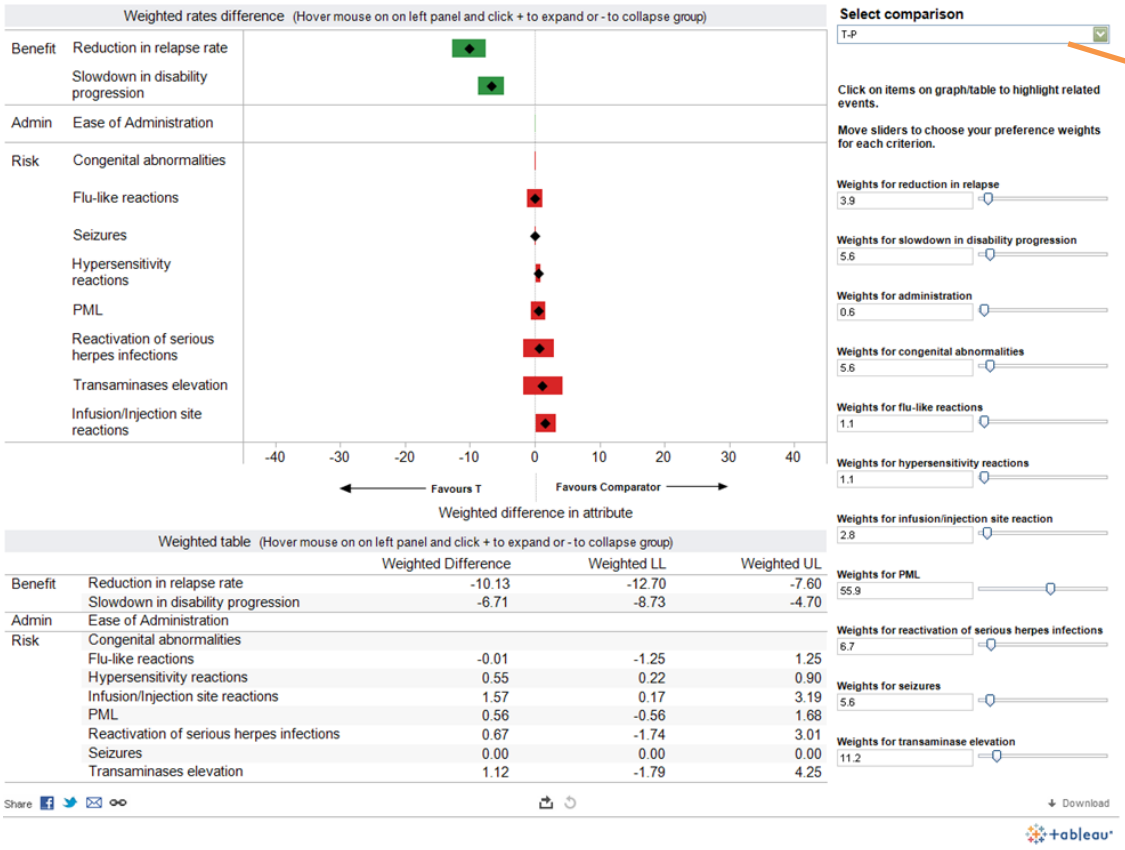
Audience: Physicians and regulators.
Communication: To communicate the benefits and risks difference when natalizumab is compared to placebo or drug by each individual criterion.
Type of plots: Forest plot and bar graph.
Annotation: Texts should be sans-serif fonts (Arial, Tahoma or Verdana), size 10+ points, colour black, and normal style
Axes: Previous axis label is fine. Arrows colour should be black and emanating from zero line. Labels for the arrow should be in black texts and comparator is kept as "comparator" for treatment blinding purpose. Weights on vertical axis to be added as later chart in bar chart form. Better explained by "weights refer to judgement provided by a group of patients asked to compare the different outcomes for their importance (high number means high importance). Sort according to importance. Rename benefit criteria to more explicit value statements: as "reduction in relapse rate" and "slowdown in disability progression" to describe benefits. "Administration" and "congenital abnormalities" are shown as missing criteria due to no rates data being available. The scale on the horizontal axis is fixed to allow for arrows but this has drawback of having values that are not very visible (software limitation).
Styles: Midpoints (mean risk difference) are represented by filled black diamonds, and the uncertainties are represented as horizontal 95% confidence intervals bar of the means. Preference values are represented by horizontal bars with similar theme to the forest plot.
Colours: Use green for benefit criteria and red for risk criteria instead of blue and yellow. Use light green for (ease of) administration criterion.
Content/storylines: - first image: Forest plot of the difference in rates of events for all criteria arranged in order of magnitude by higher level criteria (benefit/risk). Additionally, add a bar chart for corresponding elicited weights. Next image (currently unlinked) is a forest plot with the combination of weights and rates difference (individually weighted benefit-risk difference), plus a table of the values for easy reference defaulted to show results at the elicited weights combination. - filter: Radio buttons (Figure 7.7) and drop-down list (Figure 7.8) are included to filter the comparator, whether to show natalizumab compared to Anonex (T-A), Glatiramer acetate (T-C) or Placebo (T-P). - animation: Expandable/collapsible benefit-risk hierarchy. Allow users to highlight and sort all related components on the graphs. Tooltips are included to show the values of the underlying data. The weighted difference graphics have sliders for each criterion for users to select their own weights. - drill-down: Hierarchy of benefit and risk criteria.
Data structure: Variables in the data file are *T1* – numeric value of the main treatment *T2* – numeric value of the comparator *D* – the difference between T1 and T2 (numeric value of the "dot") *max* – numeric value of the upper limit of confidence intervals (the right side of the bar part on the plot) *min* – numeric value of the lower limit of confidence intervals (the left side of the bar on the plot) *W* – numeric preference weights in percentage (optional – not used in graphic) *ID* – numeric identifier of the criteria. ID is used to control the preference weight sliders, so would require more work in Tableau to relabel the sliders if any changes is made *Desc* – text identifier of the treatment comparison to use with filtering. Texts will appear "as is" in the filter box *Note* – text description of data status (optional – just as a reminder)

Figure 7.7 Forest plot with bar graph of preference weights



http://public.tableausoftware.com/views/T_Forest/ForestAndWgt
http://public.tableausoftware.com/views/T_Forest/ForestAndWgt

Figure 7.8 Forest plot with table of outcome measures



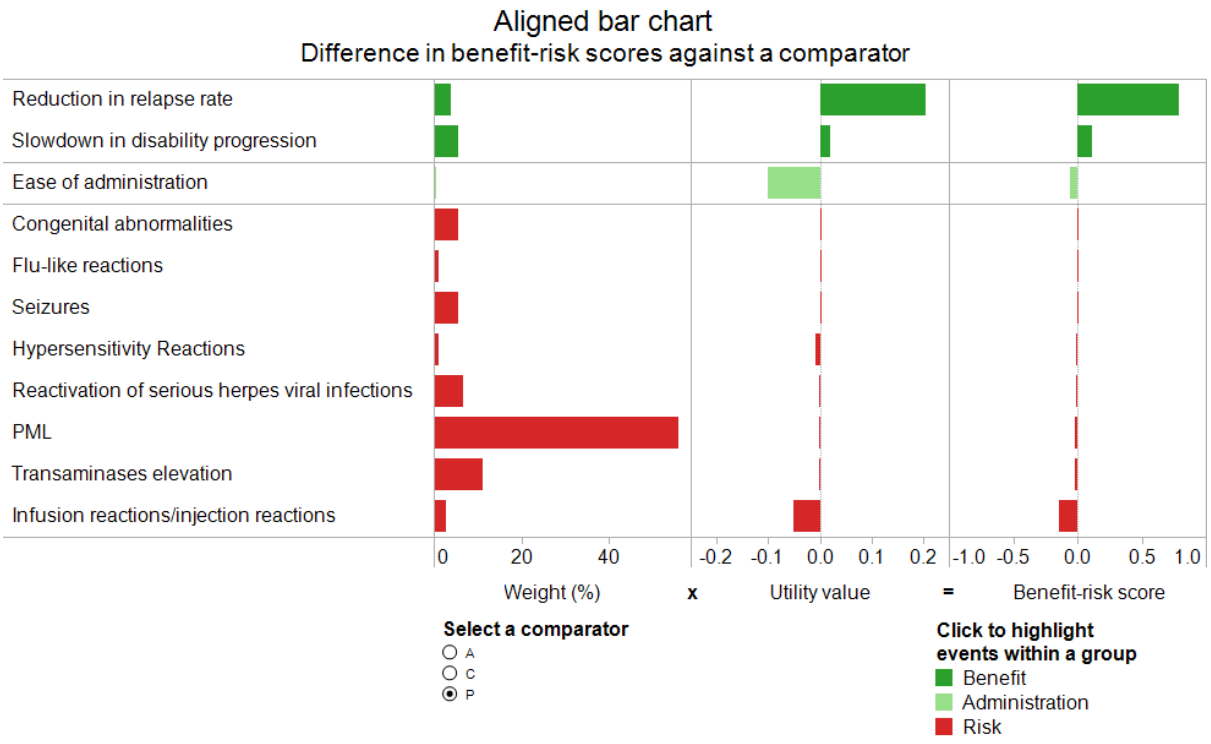
Drop-down list

http://public.tableausoftware.com/views/T_Forest/WgtNCB

7.5.4 Aligned bar chart

Audience: All (Regulators, physicians, patients, industry)
Communication: To communicate the magnitudes of benefit and risk criteria, and their weight and utility value components.
Type of plots: Aligned bar charts
Annotation: Notes are added at the bottom of the graph region to briefly explain the concepts of weights, utility value and benefit-risk score to the users. Since the length of the horizontal axis is very short on the individual graph. We also added simple text annotation in the tooltips to explain what the benefit-risk score means to each criterion (Figure 7.10 – Figure 7.12). This is done on the simple rule: if score<0 then natalizumab is worse than Placebo (P), Anovex (A) or Glatiramer acetate (C); if score>0 then natalizumab is better than Placebo (P), Anovex (A) or Glatiramer acetate (C); and if score=0 then natalizumab is about the same as Placebo (P), Anovex (A) or Glatiramer acetate (C). Better rule is required to account for uncertainty and regions of clinical equivalence.
Axes: Axes are labelled with the corresponding variable name (Weight, Utility Value, Benefit-risk score). They are labelled to show that the product of weight and utility value yielded the benefit-risk score. The scales on the axes (range) are not fixed to allow visibility since changing the comparator may also change the scale.
Styles: All criteria values/contributions are shown as bars. The lower level criteria are placed within hierarchy of higher-level criteria (benefit, risk, administration). The weight is represented as a simple bar chart, whilst the utility values and benefit-risk scores are represented as a "difference display" where a positive value is in favour of natalizumab, and a negative value is in favour of a chosen comparator.
Colours: Dark green represents benefit criteria, light green represents the ease of administration, and red represents risk criteria.
Content/storylines: - first image: Comparison with Placebo is shown - filter: Allow filter by comparator placebo (P), Beta-interferon (A), Glatiramer acetate (C) which are currently blinded as the graphic is live online. - animation: None - drill-down: None
Data structure: Variables in the data file are *Comparator* – text name of the comparator to the main treatment (here it is natalizumab). Text data will appear "as is" on the graphic *OUTCOME* – text name of the criteria. Text data will appear "as is" on the graphic *wgt* – numeric preference weights as percentage (should add up to 100%, but weight individual weight is calibrated to the total weight by comparator) *value* – numeric value of the difference between the main treatment and the comparator (e.g. Natalizumab-Placebo)

Figure 7.9 Aligned bar chart showing the weight, utility values and benefit-risk score by criterion



Notes:
1.' Weights (%) ' shows the elicited preference weights obtained from patients representatives, on a scale of 0 (least preferred) to 100 (most preferred). In the case of a benefit, 100 refers to most preferred to achieve. In the case of a risk, 100 refers to most preferred to avoid. Likewise for preference weight of 0.

2. 'Utility value' shows the difference in transformed utility value of an attribute based on the probability of occurence of the respective event (refer to XX PROTECT report for details of transformation). Utility value ranges from 0 (least favourable) to 1 (most favourable).

3. 'Benefit-risk score' shows the attribute benefit-risk score as difference in weighted utility values between T and a comparator. This is calculated as 'benefit-risk score' = 'weight' x 'utility value'.

tableau

http://public.tableausoftware.com/views/T_AlignedBar/AlignedBar

Figure 7.10 Positive judgement call in the tooltip annotation to assist user to understand the benefit-risk score

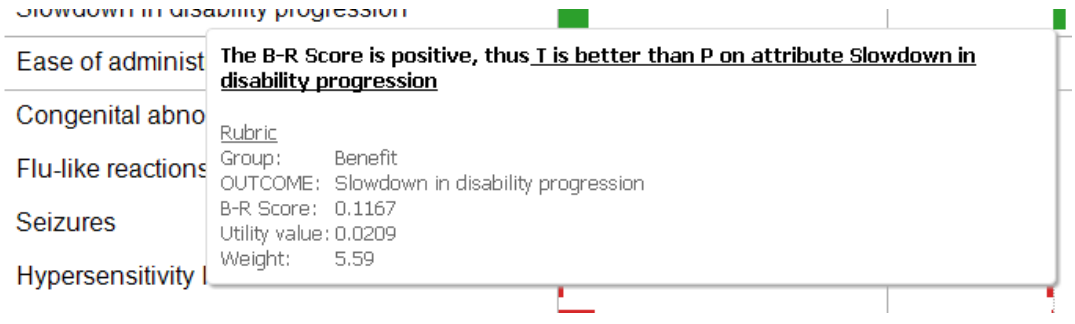


Figure 7.11 Equivalence judgement call in the tooltip annotation to assist user to understand the benefit-risk score

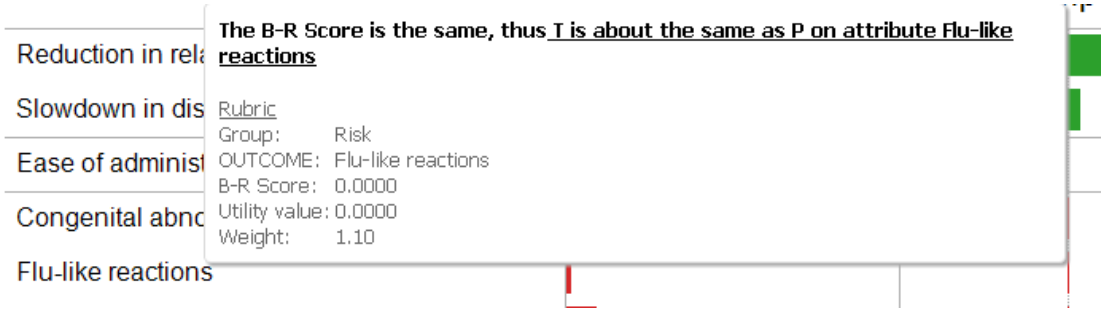
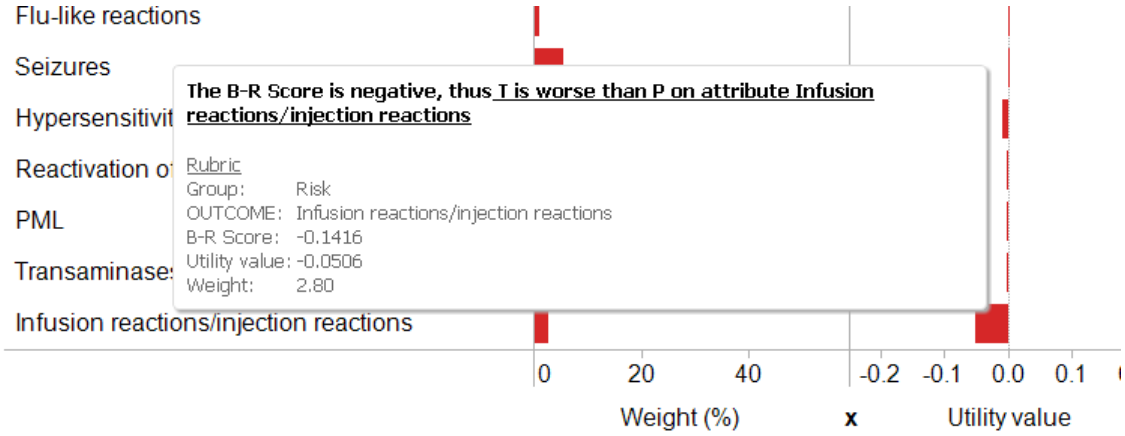


Figure 7.12 Negative judgement call in the tooltip annotation to assist user to understand the benefit-risk score



7.5.5 Waterfall plot

Audience: All (Regulators, physicians, patients, industry)
Communication: To communicate the additional contribution of a criterion to the cumulative benefit-risk score comparing natalizumab to placebo, Beta-interferon or Glatiramer acetate.
Type of plots: Waterfall plot
Annotation: Texts should be sans-serif fonts (Arial, Tahoma or Verdana), size 10+ points, colour black, and normal style. Here chosen Arial 11 points.
Axes: Based on the data, the horizontal axis is labelled as "Cumulative benefit-risk (EDSS)". The axis range currently is not fixed to allow viewable graph area but can be fixed once the actual range is known.
Styles: All criteria values/contributions are shown as bars. The lower level criteria are placed within hierarchy of higher-level criteria (benefit, risk, administration).
Colours: Use green for benefit, red for risk, orange for administration. Bar showing the overall benefit-risk score is coloured purple.
Content/storylines: - first image: A plot showing cumulative benefit-risk score against placebo as the comparator to natalizumab. The order is somewhat random. - filter: Allow filter by comparator placebo (P), Beta-interferon (A), Glatiramer acetate (C) which are currently blinded as the graphic is live online. - animation: Expandable/collapsible benefit-risk hierarchy. Allow users to highlight and sort all related components on the graphs. Tooltips are included to show the values of the underlying data. - drill-down: Allowed by hierarchy of criteria.
Data structure: Variables in the data file are *OUTCOME* – text name of the criteria. Text data will appear "as is" on the graphic *min* – numeric value of the left side of the bar (lower value) *max* – numeric value of the right side of the bar (higher value) *sise* – (numeric) difference of max-min (calculated field in the Excel datasheet) *page* – text name of the comparator to the main treatment (here it is natalizumab). Text data will appear "as is" on the graphic *grp* – text name of the high-level category to use with filtering. Text data will appear "as is" on the graphic

Figure 7.13 Waterfall plot colour-coded by natalizumab's performance against a comparator

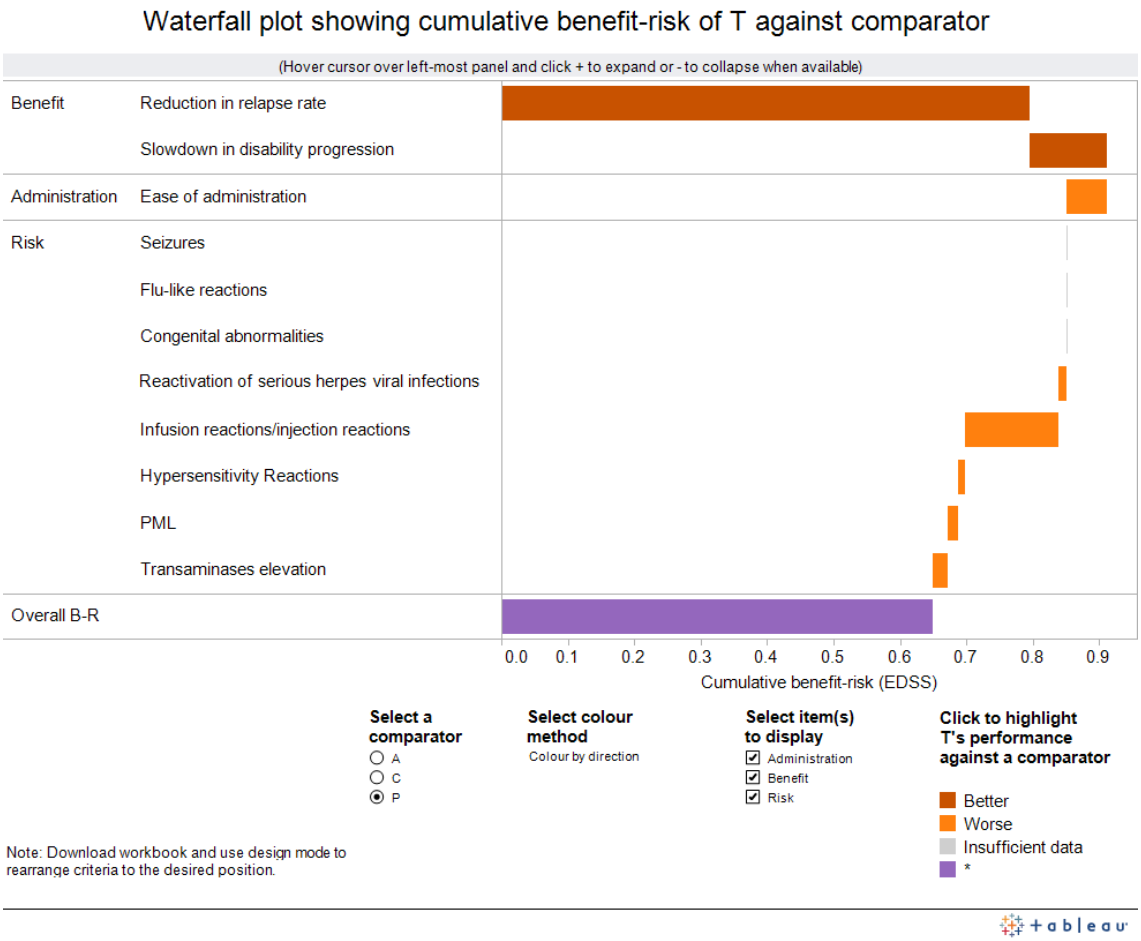
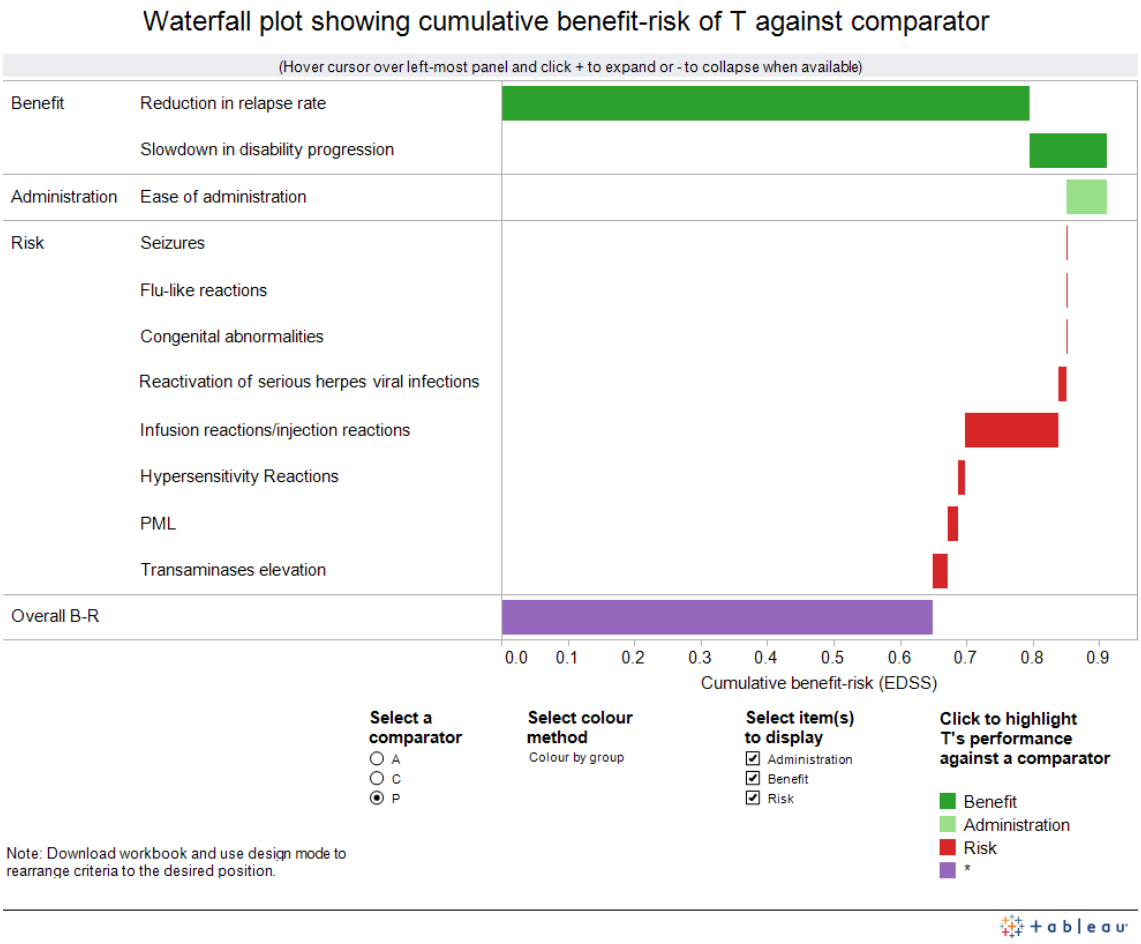
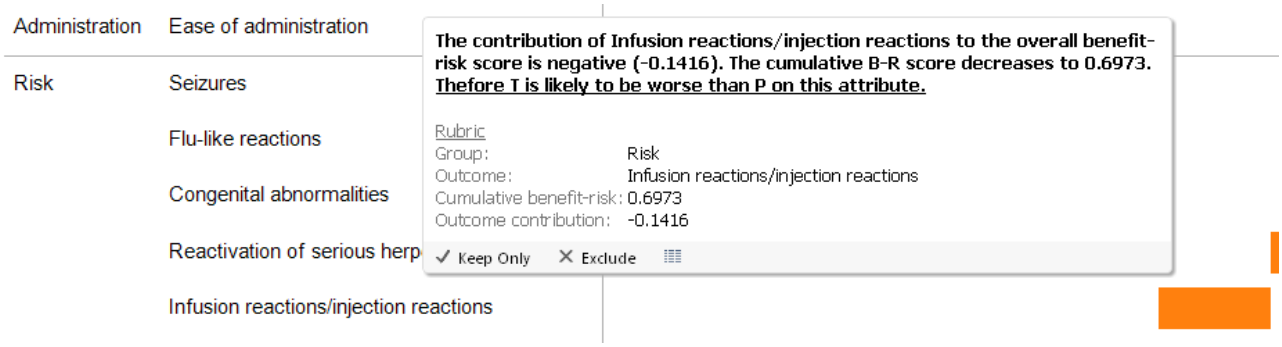


Figure 7.14 Waterfall plot colour-coded by high-level criteria (group)



http://public.tableausoftware.com/views/T_Waterfall/WaterfallRisk

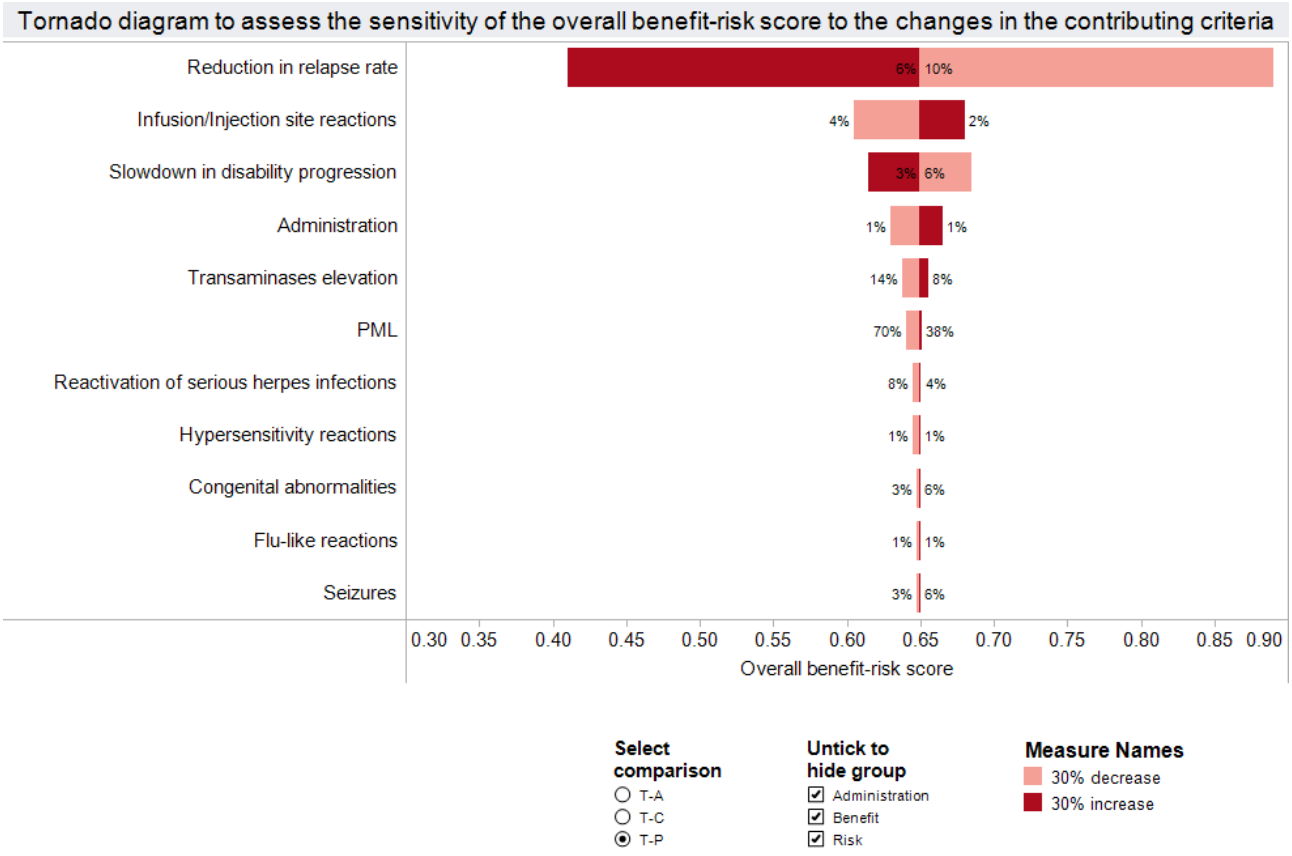
Figure 7.15 Example of negative judgement call in the tooltip annotation on "infusion/injection reactions" criterion to assist user to understand the benefit-risk score



7.5.6 Tornado diagram

Audience: Statisticians and regulators
Communication: To communicate how the changes in the natalizumab outcome measure affect the overall benefit-risk score by changing a fixed amount (10% and 90% quantiles) of the measured outcomes.
Type of plots: Tornado diagram
Annotation: Texts should be sans-serif fonts (Arial, Tahoma or Verdana), size 10+ points, colour black, and normal style. Here chosen Arial 10 points. Value labels are displayed to represent the low and high values on the bar for quick assessment. Unmeasureable outcomes, such as when there is no data or data is categorical, are labelled as "Indeterminate". The base values are not displayed but are included in the tooltips.
Axes: The horizontal axis is on continuous scale and is labelled as "Overall benefit-risk score (#relapse)".
Styles: The criteria are ordered by the magnitude of change in the overall benefit-risk score to get the tornado-like effect.
Colours: Colour-coding is done via the direction of change i.e. whether a bar refers to a lower value or higher value of the outcome measure. Colours used are of high and low saturation. Red colour scheme is chosen.
Content/storylines: - first image: Default to showing full tornado diagram comparing natalizumab against placebo. - filter: Allow filter by comparator placebo (P), Beta-interferon (A), Glatiramer acetate (C) which are currently blinded as the graphic is live online. There is also a filter to hide a particular group of criteria (Administration, Benefit or Risk). - animation: None. - drill-down: Not applicable
Data structure: Variable in the data file are *Desc* – text name of the comparator to the main treatment (here it is natalizumab). Text data will appear "as is" on the graphic *Criteria* – text name of the criteria. Text data will appear "as is" on the graphic *brmid* – numeric value of the base overall BR score *brmin* – numeric value of the minimum overall BR score with the lowest value of criterion for risk criteria and highest value for benefit criteria. *brmax* – numeric value of the maximum overall BR score with the highest value criterion for risk criteria and highest value for benefit criteria. *mid* – the base value of criterion *min* – the lowest value of criterion *max* – the highest value of criterion *Wgt* – numeric preference weight in percentage (optional – not used in graphic) *Note* – text description of data status (optional – not used in graphic)

Figure 7.16 Tornado diagram to assess the sensitivity of benefit-risk score to changes in the underlying criteria comparing natalizumab to a comparator



Notes:
1. The 30% value to analyse the sensitivity for change in score is was determined arbitrarily.
2. Preference weights incorporated into the analyses were those of patient representatives elicited through a decision conference.
3. Evidence data on the benefits and risks events were obtained from published clinical trials.

tableau

http://public.tableausoftware.com/views/T_Tornado/T_Tornado

Figure 7.17 Descriptive tooltips of the associated values and their ranked importance to the overall BR score