# IMI Work Package 5: Report 1:b:iii Benefit - Risk

# Wave 1 Case Study Report:

# Raptiva® (efalizumab)

# 04/02/2012

Alain Micaleff (MerckSerono SA)
Tornbjorn Callreus (DKMA)
Lawrence Phillips (EMA, LSE)
Diana Hughes (Pfizer)
Kimberley Hockley (Imperial College London)
Nan Wang (Imperial College London)
David Luciani (Mario Negri Institute)
On behalf of PROTECT Work Package 5 participants

| Version one dates 23 Jan 2013<br>Date of any subsequent amendments below | Person making amendments | Brief description of amendments |
|---|---|---|
| V.2.1 04/02/2013 | Larry Phillips | Various edits across document |
| 17/06/2013 | Alain Micaleff / Shahrul Mt-Isa | Appendix 9.1 expanded into two supplements to this report. |

https://eroombayer.de/eRoomReq/Files/PH-GDC-PI-SID/IMI-PROTECT/0_f9082/PROTECT WP5 report template.docx

Disclaimer: The processes described and conclusions drawn from the work presented herein relate solely to the testing of methodologies and representations for the evaluation of benefit and risk of medicines. This report neither replaces nor is intended to replace or comment on any regulatory decisions made by national regulatory agencies, nor the European Medicines Agency

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

## TABLE OF CONTENT

## Glossary and Abbreviations

| | |
|---|---|
| PASI75 | Percentage of patients achieving at least a 75% reduction of PASI at week 12 when compared to baseline. The PASI is a measure of the average redness, thickness and scaliness of the lesions (each graded on a 0-4 scale), weighted by the area of involvement. PASI range is from 0 to 72. |
| PASI50 | Percentage of patients achieving 50% reduction in baseline PASI at week 12. The PASI is a measure of the average redness, thickness and scaliness of the lesions (each graded on a 0-4 scale), weighted by the area of involvement. PASI range is from 0 to 72 |
| PGA | Percentage of patients achieving Physician's Global Assessment clear/almost clear at week12. This is a seven point scale with 7 being clear, 6 almost clear, 5 mild, 4 mild to moderate, 3 moderate, 2 moderately severe and 1 severe psoriasis. |
| OLS | Percentage of patients with Overall Lesion Severity rating of minimal or clear at FT (week 12). OLS is a static global assessment with 6 categories (clear, minimal, mild, moderate, severe and very severe) based on the characteristics of plaque elevation, scaling and erythema. |
| DLQI | The DLQI is a 10-item questionnaire that incorporates patients' assessments of itch, pain, feelings of embarrassment and self-consciousness, problems with their psoriasis treatment, and interference of their psoriasis with daily activities, relationships, and sexual activity. The DLQI scores range from 0 (no impairment) to 30 (maximal impairment). |
| AE | Adverse Event: Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment |
| ADR1 | Mild to moderate AEs with frequency "very common" in the SPC |
| ADR2 | Serious ADRs including post-marketing safety experience |
| Aseptic meningitis | A syndrome characterized by headache, neck stiffness, low grade fever, and CSF lymphocytic pleiocytosis in the absence of an acute bacterial pathogen. Viral meningitis is the most frequent cause although |

| | |
|---|---|
| | mycoplasma, and rickettsia infections; diagnostic or therapeutic procedures; neoplastic procedures; septic perimeningeal foci; and other conditions may result in this syndrome. (From Adams et al., Principles of Neurology, 6th ed, p745) |
| CMV | Cytomegalovirus |
| EBV | Epstein-Barr virus |
| Haemolytic anaemia | A form of anaemia due to haemolysis, the abnormal breakdown of red blood cells (RBCs), either intravascular or extravascular |
| ISS | Integrated Safety Summary |
| Opportunistic Infections | An opportunistic infection is an infection caused by pathogens, such as bacterial, viral, fungal or protozoan infections that usually do not cause disease in a healthy host. A compromised immune system, however, presents an "opportunity" for the pathogen to infect. |
| PML | Progressive Multifocal Leucoencephalopathy, is a rare and usually fatal viral disease that is characterized by progressive damage or inflammation of the white matter of the brain at multiple locations. It occurs almost exclusively in people with severe immune deficiency, such as transplant patients on immunosuppressive medications, patients receiving certain kinds of chemotherapy, patients receiving natalizumab (Tysabri) for multiple sclerosis, psoriasis patients on long-term efalizumab (Raptiva) or AIDS patients. It is caused by a virus, the JC virus, which is normally present and kept under control by the immune system. Immunosuppressive drugs prevent the immune system from controlling the virus. |
| PMS | Post-marketing surveillance |
| RCT | Randomized Controlled Trial |
| Serious Infections | Infections which are fatal, or life-threatening, or requires inpatient hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability, or is a significant medical event. |

# 1 Introduction and medical background

The overall objective of WP5 is to assess the feasibility of various methodologies for B-R assessment of medicinal products including the provision of usable data and information, the underpinning modeling and the presentation of the results, with a particular emphasis on visualization methods. In order to integrate these various components, it has been decided during Year 2 of the PROTECT project to progress the B-R Work Package 5 through selected case studies. The first wave of these case studies, intended to test several B-R methods and several perspectives, included Acomplia® (rimonabant), Ketek® (telithromycine), Raptiva® (efalizumab) and Tysabri® (natalizumab).

This document is the report of the efalizumab Task Force, and describes the current status of the team as of December 2011.

## 1.1 The Psoriasis disease.

Psoriasis is an autoimmune disease that appears on the skin. It occurs when the immune system mistakes the skin cells as a pathogen, and sends out faulty signals that speed up the growth cycle of skin cells. Psoriasis is not contagious. However, psoriasis has been linked to an increased risk of stroke. There are five types of psoriasis: plaque, guttate, inverse, pustular and erythrodermic. The most common form, plaque psoriasis, is commonly seen as red and white hues of scaly patches appearing on the top first layer of the epidermis. Some patients, though, have no dermatological symptoms.

The disorder is a chronic recurring condition that varies in severity, location and area, from minor localized patches to complete body coverage. Fingernails and toenails are frequently affected (psoriatic nail dystrophy) and can be seen as an isolated symptom. Psoriasis can also cause inflammation of the joints, which is known as psoriatic arthritis. Between 10-30% of all people with psoriasis also have psoriatic arthritis.

The cause of psoriasis is not fully understood, but it is believed to have a genetic component and local psoriatic changes can be triggered by an injury to the skin. Various environmental factors have been suggested as aggravating to psoriasis, including stress, withdrawal of systemic corticosteroid, as well as other environmental factors, but few have shown statistical significance. There are many treatments available, but because of its chronic recurrent nature, psoriasis is a challenge to treat. Withdrawal of corticosteroids (topical steroid cream) can aggravate the condition due to the 'rebound effect' of corticosteroids but this may be followed by cure.

Although not life-threatening in its most common plaque forms, a major feature of the disease is its social impact. Severe cases of psoriasis have been shown to affect health-related quality of life to an extent similar to the effects of other chronic diseases, such as depression, hypertension, congestive heart failure or type 2 diabetes. Depending on the severity and location of outbreaks, individuals may experience significant physical discomfort and some disability. Itching and pain can interfere with basic functions, such as self-care, walking, and sleep. Plaques on hands and feet can prevent

individuals from working at certain occupations, playing some sports, and caring for family members or a home. Plaques on the scalp can be particularly embarrassing, as flaky plaque in the hair can be mistaken for dandruff. Medical care can be costly and time-consuming, and can interfere with an employment or school schedule.

Individuals with psoriasis may also feel self-conscious about their appearance and have a poor self-image that stems from fear of public rejection and psychosexual concerns. Psychological distress can lead to significant depression and social isolation.

## 1.2 Psoriasis indication of efalizumab

Efalizumab is an anti-CD11 monoclonal antibody drug that was developed to treat moderate to severe plaque psoriasis. It was hoped that the biologic compound would overcome the serious long term toxicities of some conventional systemic therapies (e.g. methotrexate, cyclosporine). The drug was approved in September 2004 by the European Medicines Agency with the following indication: "Treatment of adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA". This limitation in the indicated patient population was qualified as "high need" psoriasis patients, and was supported by a prospectively defined subgroup in one of the pivotal RCTs based on which the Marketing Authorization was granted (ref 1)

At the time of approval, efalizumab was the only available recombinant monoclonal antibody for treatment of the disease. However several alternative biological therapies (etanercept, infliximab, adalimumab) were also granted approval for treatment of moderate to severe plaque psoriasis in "high need" patients in the subsequent months. Most of these biologicals have additional indications whilst plaque psoriasis was the only approved indication for efalizumab. Therefore the safety experience of these competitors is based on a larger exposed population including several indications.

The CHMP gave a positive opinion for the Market Authorization in September 2004 based on the following assessment:

> "In prospectively designed patient population who were not controlled by, contraindicated to or intolerant to two or more systemic therapies as judged from the patient histories of psoriasis treatment, the absolute difference between the response to efalizumab and placebo was approximately 27% for the primary endpoint (PASI 75) both in the total and the restricted populations. This efficacy data, in line with results from previous studies, indicate modest efficacy (in term of PASI 75 response rate). Nevertheless it is clinically relevant in patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or have contraindication to or are intolerant to other systemic therapies. The median time to relapse in patients who initially responded to treatment (PASI 75 at week 12) ranged from 59 to 74 days following last efalizumab dose. Therapy may be continued only in those patients who respond adequately to treatment. Re-treatment may be associated with lower or inadequate response to efalizumab than in the earlier treatment period.

> The overall exposure was 2,500 patient-years. Efalizumab appears to be safe and well tolerated; the most frequent adverse drug reactions observed were mild to moderate dose-related acute flu-like symptoms. Leukocytosis and lymphocytosis were also very common but lymphocytes returned to base line after therapy. Thrombocytopenia was uncommon but platelet count monitoring is recommended. The long term safety data are limited especially for the risk of infections, the risk of auto-immune diseases, the potential for induction suppression of humoral and cellular immunity and the risk of malignancies. Psoriasis adverse events including erythroderma or pustular psoriasis were reported (mainly in non-responders) and the possibility of a rebound mechanism cannot be ruled out. There were too limited data to exclude the potential emergence of neutralizing HAHA and related complications.
>
> In order to further assess potential risks especially of auto-immune reactions, malignancies, infections, adverse events due to anti-efalizumab antibodies and interactions, the Company will perform a Post Marketing Surveillance programme."

The marketing authorisation was later suspended in February 2009, due to the occurrence of rare but fatal cases of progressive multifocal leukoencephalopathy (PML) possibly associated with long term use. Four cases of PML attributable to efalizumab were reported; three were fatal. These PML cases occurred in addition to several previous serious safety issues leading to SPC Variations in EU and US. The Market Authorisation was finally withdrawn by the Manufacturer (MerckSerono SA) in June 2009 (ref 1)

The CHMP concludes in its scientific assessment dated January 2009 that: the efficacy of efalizumab is modest. The new safety signals that have emerged (especially PML) together with the known risk of opportunistic infections do compromise the benefit/risk ratio. Since the grant of the Marketing Authorisation, safety issues have arisen leading to the addition of a number of warnings into the SPC such as aseptic meningitis, immune-mediated haemolytic anaemia, decreased antibody development with vaccinations, interstitial pneumonitis, arthritis, erythema multiforme, inflammatory polyradiculoneuropathy (Fisher-Miller syndrome, facial palsy and Bells palsy) and severe infections, malignancies during long term use, serious fatal events such as opportunistic infections and Guillain-Barre syndrome. In addition the MAH recently notified the EMA about 3 cases of encephalopathy and 5 cases of encephalitis. Furthermore, based on a comparative evaluation of serious adverse events it appears that efalizumab has an unfavourable safety profile as compared to the other biologicals with respect to fatal reports of infections, neoplasms, and neurological disorders.

## 1.3 Epidemiology of the disease

The incidence of psoriasis was estimated in a population-based study performed in Rochester (Minnesota) during a 4-year period (1980-83) on a predominantly caucasian population (Bell, 1991). The rate adjusted by age and sex to the 1980 US white population was 60.4/100,000.

In Europe, a recent study performed in a large primary care database (GPRD) (Huerta, 2006 – poster) reported an incidence rate of 15/10,000.

In a literature review by Radulescu (2006), the overall prevalence of psoriasis was between 1.1% and 2.8% of the adult population.

UK (GPRD): 1.5% over a 5-year period (Gelfand, Arch Dermatol 2005). Twenty four percent received 5 or more prescriptions for psoriasis in the first year after the first GPRD record of psoriasis. Systemic agents were used by 2.3% of the psoriasis patients.

US: lifetime prevalence of self-reported psoriasis: 2.2% (Stern 2004). The current extent of disease was 3-10 palms and more than 10 palms in 11% and 3% of the cases.

Psoriasis *per se* does not entail substantial mortality but it is associated with risk factors influencing mortality.

In a Finnish population-based cohort of patients with a hospital admission for psoriasis between 1973 and 1984 was followed up by linkage to the Cause of Death Registry through 1995. The all-cause standardized mortality ratio (SMR) was 1.62 and 1.54 in men and women, respectively. Alcohol- and smoking-related causes were major causes for the excess mortality. (Poikolainen 1999).

Mallbris et al (2004) used a cohort of nearly 9000 patients hospitalized for psoriasis. The all-cause SMR overall risk among those admitted at least once was increased by 50%.

The mean age at date of onset is 37.4 and 34.5 years for men and women, respectively. Psoriasis is slightly more prevalent in men. Incidence increases more or less steadily with age up to the seventh decade of life.

The psoriasis disease is associated with a number of co-morbidities which may be confounded with adverse effects of various treatments used in that indication. Those of clinical importance:

- Cardiovascular risk factors are associated with psoriasis, especially in its severe forms. Extensive reviews on the subject conclude that chronic inflammation in psoriasis has an unfavorable effect on the CV risk profile.
- Patients with severe psoriasis have a slightly increased risk of solid tumors of approximately 30%-50% when NMSC and lymphoproliferative disease are not taken into account. The increased risk is mainly attributable to smoking-related cancers (Olsen 1992, Hannuksela-Svahn 2000). The risk for NMSC in patients with severe psoriasis is increased by 2.5- to 4-fold compared to the general population
- Several epidemiological studies have investigated the association between psoriasis and lymphoma with different conclusions: some studies have suggested that there was no link between the two conditions whereas other studies indicated an increase in the risk of NHL

and HD of two- to eight-fold among patients with severe psoriasis. An increased risk is also described in patients with mild psoriasis (unexposed to systemic medication) (Gelfand et al 2003, Margolis 2001). The baseline risk of LPDs conferred by psoriasis should be taken into account in the evaluation of the carcinogenic potential of drugs.

## 1.4 Background epidemiology of identified and potential risks in psoriasis patients

For some of the identified or potential risks, namely, thrombocytopenia, hemolytic anemia, inflammatory polyradiculoneuropathy and facial palsy, PML, tuberculosis, there is no evidence that the condition is associated with psoriasis, therefore the background epidemiology is assumed to be the same as for the general population.

However the background epidemiology of several identified risks of efalizumab (e.g. opportunistic infections, aseptic meningitis, immune hemolytic anemia, inflammatory polyradiculoneuropathies, interstitial lung diseases) is not known in a psoriasis population, which limits the measure of Risk Difference and Relative Risk for these outcomes. This was a limitation in the use of B-R methods which require such measures.

## 1.5 Rationale for proposing this drug as an example of a Benefit-Risk quantification and representation

- Market Authorisation was controversial from the start (negative opinion from Rapporteur, comparison only with placebo and not with other standard active treatments).
- Efficacy was qualified "modest" upon review of all clinical trials at the time of the Market Approval as well as in 2009 at the time of re-evaluation by CHMP.
- The unique indication was in a non life-threatening indication (psoriasis), but with important patient utilities because of the social impact of the severe forms of psoriasis (but the drug was also possibly prescribed in less severe forms).
- There were competitors in 2009 both among other biotechnology products and small molecules, offering therapeutic alternatives.
- Continuous accrual of safety information lead to 8 major labelling changes in EU SPC for safety reasons over 4 years.
- Final emerging safety issue was a non predictable, non preventable, hardly treatable, potentially fatal AE (Progressive Multifocal Leucoencephalopathy, 4 cases).

# 2 Aim and objectives

The European Commission requested the CHMP in January 2009 to assess the emerging safety concerns related to efalizumab and their impact on its B-R, and to give its opinion on measures necessary to ensure the safe and effective use of efalizumab and on whether the Marketing Authorisation for this product should be maintained, varied, suspended or withdrawn.
The objective of the efalizumab Task Force is to replicate the decision made by the CHMP in February 2009 using the same data which were available to regulators at this time, but applying descriptive frameworks, quantitative models and their graphical representation in order to test the relevance of such models in regulatory decision making.

# 3 Methods

## 3.1 Justification for selection of B-R approaches

PROTECT Work Stream B has recommended 13 approaches to be tested in the first wave of case studies (efalizumab Raptiva®, rimonabant Acomplia®, natalizumab Tysabri® and telithromycine Ketek®). Not all of these 13 methods are relevant for the efalizumab case study; nor is it the intent of the efalizumab Task Force to test all of these methods, mainly for time constraint reasons.

**Table 1 Benefit-risk approaches included for testing in efalizumab case study**

| Approach | Justification |
|---|---|
| **PrOACT-URL** | PrOACT-URL is a qualitative framework which is a convenient initial preparatory stage to application of an MCDA analysis application. |
| **BRAT** | BRAT is a framework supported by a set of guidelines and a tool that allows data to be structured for decision making. It allows for a comparison with the PrOACT-URL framework. It does not include any formal Benefit-Risk integration providing a final score, as the actual benefit risk decision is outside of the framework. |
| **MCDA** | MCDA is a method used to integrate multiple benefit and risk-criteria. It is one of the more complex methods. MCDA naturally leads to a quantitative benefit-risk balance |
| **Benefit-Risk Ratio (BBR)** | BRR is conceptually simple and general. The concept of taking the ratio of the magnitude of benefits to risks is tested with the BRAT framework, using the |

most prominent Benefit and the most prominent Risk.

**Table 2 Benefit-risk approaches excluded from testing in efalizumab case study**

| Approach | Justification |
|---|---|
| **1. NNT/NNH** | Similar to BBR, NNT/NNH was initially designed for one benefit and one risk analysis with binary and proportion endpoints (response rate, ADR incidence etc). Later there are generalizations to combine multiple risks with relative utility (benefit can be treated similarly), but theoretical violations of utility theory precluded its use here. |
| **2. Population Impact Numbers** | It is not meaningful to apply a modified NEPP-approach (Number of Events Prevented in your Population) to very rare events with a close to zero baseline risk in the unexposed. |
| **3. QALY (Q-TWiST, INHB)** | With the summary data in EPAR, QALY is not directly applicable. A benefit-risk analysis can be performed by QALY or Q-TWiST, only if QALY or Q-TWiST are available in the trials or studies or derivable from the available information |
| **4. Probability Simulation (PSM)** | The purpose of probability simulation is to assess the uncertainty in benefit-risk criteria and how the uncertainty affects the results of benefit-risk analysis. Probability simulation can go together with any model analysis, including MCDA. |
| **5. SMAA** | SMAA is an extension of MCDA using probability distribution instead of statistical data summaries for each criterion. Uncertainty about criterion weights can also be described with probability distributions. The realization of SMAA is MCDA plus simulation. |

## 3.2 Overview and analysis approach

The efalizumab Task Force started with the development of a comprehensive framework, pre-requisite for any subsequent use of other quantitative methods. PrOACT-URL was chosen as an appropriate tool and was developed with sufficient details as to allow for the subsequent identification of source documents, relevant medical outcomes and available measures. This document served as a guidance for the identification of the source documents intended to populate the Effects Table with relevant data pertaining to Favourable and Unfavourable Effects of efalizumab. In doing so, consideration was made about the requirement for using only "publicly available" documentation. This documentation finally consisted mainly in the efalizumab EPAR, the efalizumab European SPC (ref1), and published epidemiology references. The last PSUR before MA suspension (PSUR 10) was also used (available to Assessors at the time of the 2009 re-evaluation), as it included most of the cumulative data pertaining to rare events reported in post-marketing period.

In parallel, two members of the team (KH and DH) applied a modified PhRMA BRAT framework, using the same data set as used for the MCDA analysis; using the visualization software available with the PhRMA BRAT framework, a value tree, data source table and two visualizations, the benefit risk summary table and forest plot were constructed. .

Subsequently, an MCDA analysis, using a commercial software (Hiview3) was conducted, including a Decision Conference involving all Task Force participants and an external safety expert.

Finally a benefit-risk ratio quantification was performed by one member of the team using the most prominent Benefit and the most prominent Risk of efalizumab as analysed in the BRAT framework.

# 4 Evidence data

The data sources and the choice of outcomes and measures was lead by the decision context documented in the PrOACT-URL frame work. Therefore this decision context was similar for the two methods tested (BRAT and MCDA).

| **Table 1: Define decision context** | |
|---|---|
| Objective | To evaluate the benefit-risk balance of efalizumab with the use of safety and efficacy data obtained from clinical trials and cumulative post-marketing safety information on 2009, in order to examine the impact of utilizing a structured benefit-risk assessment. |
| Drug | Raptiva® (efalizumab) |
| Dose | An initial single dose of 0.7 mg/kg body weight is given followed by weekly injections of 1.0 mg/kg body weight, subcutaneously |
| Drug class | Monoclonal antibody |
| Formulation | All (e.g. GNE SC, XOMA SC) |
| Indication under consideration | Moderate to severe plaque psoriasis |
| Intended patient population of interest (including contraindications to treatment and baseline disease characteristics) | "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA |
| Comparator(s) | Placebo as reported in RCTs in the 2004 submission |
| Time horizon (for outcomes to occur), i.e. time frame for treatment and for follow-up for relevant clinical outcomes | 12 weeks for PASI 75 (efficacy), 3 years for PML (safety) |
| Decision-maker perspective (e.g. regulator, sponsor, patient, physician) | Regulator |

## 4.1 Objective data

Most of the Favorable Effects were extracted from the EPAR and consisted in the primary and secondary endpoints of 5 pivotal randomized controlled clinical trials (4 US trials from partner Genentech and 1 European collaborative trial which included a subpopulation of "high need" patients), forming the basis for the claimed and finally approved indication.

The Unfavorable Effects consisted in the safety information collected during the development phase and included in the Integrated Safety Summary. Additionally, all relevant adverse effects which were collected and analyzed in the post-marketing phase were added to the Unfavorable Effects of the Effects Table, in order to document the accumulation of safety information following the market

approval of efalizumab worldwide in 2004. Exact figures for these post-marketing data were extracted from the PSUR No 10 which was the last one before CHMP opinion of February 2009 and includes cumulative data. PSURs are not "publicly available" documents per se, however these safety data were extensively available to regulators who had to make a decision in Feb 2009, and most of the emerging safety risks had been included in the SPC over the post-marketing years from 2004 to 2009. Based on the data, the efalizumab Task Force finally agreed with the build-up of an Effects Table (see Section 3 Effects table: Criteria Definitions and Effects of Placebo and Raptiva 1mg/kg/wk in the "Supplement to Wave 1 case study report Efalizumab, Feb 2013") including chosen criteria of Favorable Effects and Unfavorable Effects.

The fine-tuning of the Effects Table differed in the subgroup using the MCDA method and the subgroup using the BRAT framework, as time prohibited the due diligence required to ascertain comparative outcome measures or conduct the necessary data transformation required to effectively use the PhRMA BRAT software.

For the BRR analysis, an additional document including the evaluation of the exposed population in the post-marketing period using sales data, has been used. This source document is also included in the PrOACT-URL document.

## 4.2 Subjective data

For the subgroup using the MCDA approach, a group discussion on the preferences related to the above mentioned criteria took place at a "Decision Conference" held on 1st December 2011, which was attended by most of the Task Force and an efalizumab Safety expert. Despite some attempts to obtain the participation of roles representing the chosen perspective (i.e. regulators) this was not possible, and finally it was agreed that the Task Force itself would be a surrogate and play the role of such a consensus meeting. Of note the efalizumab Task Force includes a Regulator, 3 MDs and 2 biostatisticians.

For the subgroup using the BRAT framework, it was decided not to apply formal weightings to convey outcome importance, and to understand what limitations there are, if any, to determining benefit/risk in the absence of delineating outcome importance.

# 5 Results

## 5.1 PrOACT-URL

PrOACT-URL (ref 2) is a generic framework for decision making, as explained in Hammond JS, Keeney RL, Raiffa H, *Smart Choices: A Practical Guide to making Better Decisions*, Boston, MA: Harvard Business School Press; 1999. This is a 12-step process which can be used in several settings, but has recently been adapted for modelling benefit-risk of medicinal products, e.g. in the EMA's Benefit-Risk Project.

PrOACT-URL stands for: **Pr**oblem, **O**bjectives, **A**lternatives, **C**onsequences, **T**rade-offs, **U**ncertainty, **R**isk tolerance and **L**inked decisions.

A detailed description of each of these steps applied to the efalizumab case study has been developed and adopted by the Task Force. The format of the document includes an additional column used for the identification of the information sources to be used for the formal assessment of the Benefit-Risk balance of the drug, and from which objective data used in subsequent models are to be extracted. The pre-requisite of using only "publicly available" data has been complied with, although the case study scenario chosen by the Task Force was the regulator's perspective, keeping in mind that regulators may have access to a larger scope of documents from the Market Authorization Holder such as PSURs for marketed products or submission dossiers.

The PrOACT-URL framework does not provide any formal Benefit-Risk evaluation per se, but is a convenient stepwise process ensuring that all aspects of a case study are addressed exhaustively.

**Case Study Report: Raptiva (efalizumab)**
**as prepared according to the IMI-PROTECT Work Package 5, Work Group D guidelines**

This Guideline is based on PROACT-URL, a generic framework for decision making, as explained in Hammond JS, Keeney RL, Raiffa H, *Smart Choices: A Practical Guide to making Better Decisions*, Boston, MA: Harvard Business School Press; 1999.

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| **Pr**OBLEM<br>1. Determine the nature of the problem and its context. | **1a. Medicinal product**: The medicinal product is Raptiva (Efalizumab). Marketed biological entity. Is a recombinant, humanized IgG1 monoclonal antibody that targets CD11a, the α-subunit of leukocyte function associated antigen 1 (LFA-1). Mechanism of action may lead to inhibition of leukocyte migration, similarly to natalizumab.<br><br>**1b. Indication(s) for use**: efalizumab is indicated in the treatment of "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA. The duration of initial therapy is 12 weeks. Therapy may be continued only in patients who responded to treatment (PGA good or better).<br><br>Together, the clinical pharmacology of efalizumab and the safety and efficacy data (including 2 phase 3 studies with 1.0 mg/kg/week and 2.0 mg/kg/week) support the selection of 1.0 mg/kg/week SC as the optimal dose for efalizumab. (EPAR scientific discussion) | EPAR: EU authorisation on 20th September 2004. Suspended Feb 2009, withdrawn June 2009; |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| | **1c. The therapeutic area and disease epidemiology:** Moderate to severe chronic plaque discoid psoriasis. Psoriasis is a common chronic, squamous dermatosis with polygenic inheritance and a fluctuating course. Principal histological findings are Munro micro abscesses and spongiform pustules; also seen are rounded, circumscribed, erythematous, dry, scaling patches of various sizes, covered by greyish white or silvery white, umbilicated and lamellar scales, usually on extensor surfaces, nails, scalp, genitalia and the lumbosacral region.<br><br>**1d. The unmet medical need**: At the time of initial Market Authorisation, there are well established systemic treatments (cyclosporine, methothrexate, PUVA) all of which with serious Adverse Effects (but B-R of the drugs is well established for a long time). At the time of the re-evaluation of the B-R of efalizumab (Jan 2009) there are more recent alternative therapies (biologic treatments for moderate to severe psoriasis in "high need" adult patients e.g. adalimubab, etanercept, infliximab, ustekinumab. with established efficacy but long term safety still uncertain in the psoriasis indication (although with longer experience in other indications such as RA)<br><br>**Severity of condition**: Psoriasis is a chronic disease, leading in its severe forms to a significant social disability impacting both professional and social life. Although psoriasis is a serious disease, with potential severe negative impact on the patient's social life, it is not a life-threatening disease apart from rare erythrodermic forms which were excluded from the clinical trials population and was not part of the approved indication (nor were pustular forms of the disease and psoriatic arthritis)<br><br>**Affected population**: "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA | Standard Text Books<br><br>efalizumab RMP update Nov 2008 pages 30-40<br><br>CHMP Opinion EMEA/CHMP/3552/2009;<br><br>Serono internal data: Serono analysis of patients treated with efalizumab after previous treatment with anti-TNF (26 Jan 2009)<br><br>Rapporteurs' Final Assessement Report EMEA/H/C/00542 |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| 2. Frame the problem. | **Patient concerns**: impact on quality of life, physical appearance and social functioning<br><br>**Physician concerns**: chronic and incurable with unpredictable flare ups, interested in long term efficacy<br><br>**Time frame for health outcomes**: 12 weeks for PASI 75 (efficacy/favourable effects)), 3 years for PML (safety/unfavourable effects). PASI 75 (primary endpoint) is a 75% reduction of the PASI score at week 12.<br><br>**1e. What is to be decided**: Re-evaluation of benefit-risk of efalizumab was prompted by incidence of emerging adverse events in the post-marketing period, i.e. presentation of PML (Progressive Multifocal Leucoencephalopathy) in addition to other serious risks (cardiotoxicity, neurotoxicity, serious infections including tuberculosis). The question to be addressed is: are there in January 2009 any risk minimisation measures which could be rapidly implemented, thus maintaining the B-R balance of the drug as positive? If not, should the Market Authorisation be suspended/revoked?<br><br>**By whom**: the Case study takes the regulator's perspective (1$^{ST}$ step of the efalizumab Task Force); next perspective to be addressed is the psoriasis patient's perspective, given the significant social impact of the severe forms of the disease.<br><br>**When**: 16$^{th}$ January 2009. Experts believed the margin of benefits over risks had narrowed since approval, i.e. modest efficacy and increased risks. The European Commission initiated a procedure under Article 20 of Regulation (EC) No 726/2004 and requested the Committee to assess the above concerns and its impact on the benefit/risk balance for efalizumab, and to give its opinion on measures necessary to ensure the safe and effective use of efalizumab and on **whether the marketing authorisation for this product should be maintained, varied, suspended or revoked**. The CHMP also took advice from the Scientific Advisory Group before making a | Marketing authorisation, pivotal studies<br><br>efalizumab RMP update Nov 2008<br><br>Responses of the Scientific Advisory Group CNS to the CHMP list of questions on efalizumab 7 Jan 2009. EMA/24463/2009 |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| | decision.<br><br>The efalizumab case study intends to replicate the decision made by the CHMP in February 2009, but using various frameworks and quantitative models.<br><br>**2a. Problem of uncertainty, multiple conflicting objectives, combination of the two, or something else?**<br>The 4 PML cases are not strongly confounded. The positive diagnosis is serologically confirmed in 3/4. There is no alternative diagnosis.<br>The uncertainty relates mainly on the relationship between duration of treatment (time on exposure) and the occurrence of PML. The impact is on the possible risk minimization measure if this had been confirmed.<br>In addition to the PML risk (potentially fatal Adverse Effect), some other risks emerged during post-marketing period.<br>Risk has increased, documented with several SPC amendments over the 4 years marketing.<br>Long term treatment: some studies (ACD2058g)included a retreatment period (RT) or extended treatment (ET); there were 2 observation periods without treatment: Observation period (OB) and Follow-up (FU); ACD2059g included only 3 periods (FT, ET and FU); the results suggest that patients not responding within 3 months will be less likely to respond to prolonged treatment for another 3 months.<br>In total, data from extended treatment (more than 12 weeks) have been obtained from 4,311 patients in open label uncontrolled studies. Over 600 patients have been treated for more than 1 year including 166 patients treated for more than 2 years and up to 3 years.<br>**2b. The factors to be considered in solving the problem:**<br>**Study design:** no direct comparison with any systemic treatment (standard treatments or new biological). Topical symptomatic treatment was allowed as per investigator in all RCTs.<br>**Adequacy of data sources:** Efficacy data was obtained from 5 double blind, placebo controlled Phase III clinical trials designed to evaluate efficacy of efalizumab as a systemic monotherapy. | Scientific Conclusions<br><br>EMEA/H/C/000542/A20/0028<br><br>EMA/CHMP/3552/2009<br><br>CHMP opinion<br>EMEA/CHMP/3552/2009;<br><br>Rapporteurs' Final Assessement Report |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| | Safety data is obtained via the number of adverse event (AE) reports received in post-marketing setting by spontaneous sources (health care professionals, literature, regulatory authorities, etc.) Safety data is based on reported events and so can potentially under represent the number of events. This may be due to poor reporting and sensitivity, and there may be an insufficient timeframe to allow for development of adverse events post long term exposure to efalizumab. However, underreporting of PML is likely to be minimal due to widely circulated documentation to physicians warning the risk of PML.<br><br>**Disease epidemiology**: efalizumab is indicated in the treatment of "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA. It needs to be considered how important/essential it is that efalizumab is available to "high need" patients where other medications may not have worked.<br><br>**Presence of alternative treatments:** In September 2004, 2 biologic medicines (i.e., etanercept and) were approved in the EU for the treatment of plaque psoriasis. Subsequently, infliximab was approved for this indication in September 2005, followed by adalimumab in December 2007. All 4 biologic therapies licensed in the EU are indicated for adult psoriasis. PML cases have been reported with some of these biologicals, but not in their psoriasis indication. | EMEA/H/C/00542<br><br>PSURs and SPC Variations<br><br>efalizumab RMP update Nov 2008 pages 30-40 |
| **O**BJECTIVES<br>3. Establish objectives that indicate the overall purposes to be achieved.<br><br>4. Identify: | **3. The aim**: The aim is to evaluate the benefit-risk balance of efalizumab with the use of safety and efficacy data obtained from clinical trials and cumulative post-marketing safety information, from a regulator's perspective and using a quantitative method (MCDA) in a first step (other methods to be tested in a later stage of the efalizumab Task Force). BRAT framework will also be developed in the first step of this Case study. | |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| a) favorable effects b) unfavorable effects | **4a. Favourable effects (i.e. efficacy):** The primary efficacy endpoint is the proportion of subjects with a 75% or more improvement from baseline in the PASI score (PASI75). This endpoint is strongly recommended in conjunction with a validated standardised global score (e.g. PGA) in the EMA GUIDELINE ON CLINICAL INVESTIGATION OF MEDICINAL PRODUCTS INDICATED FOR THE TREATMENT OF PSORIASIS**.** Five pivotal clinical studies evaluating efficacy of efalizumab in moderate to severe psoriasis primarily as systemic monotherapy were submitted (ACD2058g, ACD2059g, ACD2390g, ACD2600g and IMP24011). These studies were double blind, placebo-controlled Phase III trials. In total 2714 patients received efalizumab subcutaneously (SC). These trials with efalizumab all had similar study design. In addition study 24011 had a prospectively defined "high need" population (patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA.)<br><br>The inclusion and exclusion criteria were comparable. The main inclusion criteria were a minimum Psoriasis Area and Severity Index (PASI) score of 12.0 at screening, a plaque psoriasis covering ≥10% of total body surface area (BSA) and a need for systemic treatment.<br><br>Other outcomes to be considered include PGA (percentage of patients achieving Physician's Global Assessment clear/almost clear at week 12), OLS (percentage of subjects with Overall Lesion Severity (OLS) rating of Minimal or Clear at week 12).<br><br>In some studies (ACD 2058g, ACD 2059g, and ACD 2390g) additional endpoints included mean improvement in DLQI (dermatology life quality index) and mean improvement in the frequency and severity subscales of Psoriasis Symptom Assessment (PSA).<br><br>In study 24011, an additional endpoint was PASI 50 (proportion of subjects with a 50% | CHMP Assessment Report EMEA/H/542/A20/28 (See Section 1 of the "Supplement 1 to Wave 1 case study report Efalizumab" http://www.imi-protect.eu/benefitsRep.shtml); Market Authorization/EPAR<br><br>CHMP Assessment Report EMEA/H/542/A20/28 (See See Section 1 of the "Supplement 1 to Wave 1 case study report Efalizumab" http://www.imi- |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| | improvement from baseline in the PASI score (partial responders).<br><br>**4b. Unfavourable effects (i.e. safety):** adverse events reported to be associated or caused by efalizumab (spontaneously reported Adverse Effects are deemed to be causally related to the drug per reporter).<br><br>Safety issues added to the SPC or strengthened warnings since the initial MAA of efalizumab in the EU are as follows: aseptic meningitis, (opportunistic) infections including tuberculosis, immune mediated haemolytic anaemia, decreased antibody development with vaccinations, interstitial pneumonitis, arthritis, erythema multiforme, inflammatory polyradiculoneuropathy including Guillain Barré like syndrome and Miller Fisher syndrome, facial palsy and Bells palsy during long-term use, severe infections and malignancies, PML.<br><br>Other unfavourable effects may include overall incidence of AEs per SOC in Clinical Trials at week 12.<br><br>At the time of the CHMP assessment report, the efalizumab worldwide exposed population was estimated 47,000 patient-years. An evaluation of the exposed population per duration of exposure is available (e.g. estimated number of patients exposed to efalizumab for more than 3 years ranges from 2,236 to 3,832) | protect.eu/benefitsRep.shtml, and additional notes for summary);<br>PSURs and SPC variations (See Section 1 of the "Supplement 1 to Wave 1 case study report Efalizumab" http://www.imi-protect.eu/benefitsRep.shtml, and additional notes for summary) |
| **A**LTERNATIVES<br>5. Identify the options to be evaluated against the criteria. | **5a. Pre-approval**: N/A<br>**5b. Post-approval**:<br>• do nothing, if the B-R assessment is still positive<br>• limit duration, to 2 years (proposed by MAH based on the observed delay of onset of the 4 reported cases of PML) | N/A<br>CHMP Opinion EMA/CHMP/3552/2009 |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| | • Limit duration AND restrict indication to a subset of patients where B-R would still be positive<br>• Suspend/revoke Market Authorization. | |
| **C**ONSEQUENCES<br>6. Describe how the alternatives perform for each of the criteria, i.e., the magnitude and desirability of favorable effects, the severity of unfavorable effects, and the incidence of all effects. | **Alternative: Do nothing**: implies that B-R balance still considered positive by Rapporteur and CHMP using MCDA quantitative model based on above data.<br>**Alternative: Restrictions:**<br>• (i) 2 year treatment duration limitation: no available data; would require a prospective study; no guidance for transition to alternative treatment<br>• (ii) Target population change; however the indication in EU is already restricted to the defined "high need" population.<br>• (iii) Suspension/revocation of MA: dose tapering? risk of rebound effect (rare erythrodermic forms reported upon treatment withdrawal); transition to alternative treatment (not documented, no available data nor guidance). Drug Recall Worldwide in case of revocation of MA in EU and US. | (i) Serono internal document: Risk of PML: analysis of incidence and risk reduction;<br>(ii) no efficacy and safety data, no subgroup analysis. |
| **T**RADE-OFFS<br>7. Assess the balance between favorable and unfavorable effects. | **Judgment that was made about the benefit-risk balance**: Negative Benefit-Risk Balance, voted by CHMP (20 out of 31). B-R assessment to be reiterated using the same data but with a MCDA quantitative method. | CHMP Opinion EMA/CHMP/3552/2009 |
| | | |
| **U**NCERTAINTY<br>8. Report the | **Efficacy**: Uncertainty on the extent of off-label use in patients with less severe conditions, | Efficacy: no source data on off label |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| uncertainty associated with the favorable and unfavorable effects.<br><br>9. Consider how the balance between favorable and unfavorable effects is affected by considering the uncertainty associated with the effects. | decreasing the benefit part of the balance.<br>No direct comparison with any other systemic treatment, neither standard (cyclosporine, methotrexate, PUVA) nor biologicals. Assessors of B-R in Jan 2009 had indirect comparison with results of RCT for new biologicals.<br><br>**Safety**: Uncertainty on the shape of the risk function of PML over time (probably not linear), based on only 4 cases. No true incidence but only reporting rate, although under-reporting is unlikely or very limited due to large communication of this risk to patients and prescribers An internal document provides the patient exposure per duration of treatment based on Sales data.<br><br>**The extent to which the benefit-risk balance in step 7 is reduced by considering all sources of uncertainty, to provide a benefit-risk balance**:<br>Whichever the uncertainty on efficacy and safety data, all scenarios would decrease the benefit risk balance (underestimated risk, overestimated benefit). If all deterministic measures (derived from measures of central tendency on all the criteria) were set to the favourable limits of their confidence intervals, then, clearly, the B-R ratio would improve. However, considering the full range of uncertainty usually leads to a less favourable B-R balance. Thresholds are not considered in multi-criteria decision analysis because these models just compare the benefit-risk balances of the alternatives. Decisions based on single criteria can only be justified if the entire weight of 100% is assigned to that one criterion. | use. Limited post-marketing studies.<br><br>Safety: Serono internal document: Risk of PML: analysis of incidence and risk reduction<br><br>No source data on the under-reporting rate of various AEs (possibly minimal on the major PML risk) |
| **R**ISK TOLERANCE<br>10. Judge the relative importance of the decision maker's risk attitude for this | **10. Any considerations that could or should affect the decision maker's attitude toward risk for this product (e.g., orphan drug status, special population, great medical need, risk management plan)**:<br>• Initial MA in 2004 was already controversial (no consensus between Rapporteur and co-Rapporteur)<br>• In January 2009, medical need is covered by several other therapeutic options, and | CHMP Opinion and grounds for decision. EMA/CHMP/3552/2009 |

| STEP | DESCRIBE | DATA SOURCES |
|---|---|---|
| medicinal product.<br><br>11. Report how this affected the balance reported in step 9. | efalizumab has modest efficacy when compared to alternative treatments (indirect comparison with similar endpoints from RCT with new biological)<br>• Psoriasis is not a life-threatening disease though it may have a serious impact on social and professional life<br>• Risk Management Plan with no obvious risk minimization measures which could be easily and quickly implemented (sub population?, limitation of treatment to 2 years).<br><br>**11. The basis for the decision maker's decision as to how tolerable the benefit-risk balance is judged to be (taking into account stakeholders' views of risk?):**<br>Safety Advisory Group (SAG, consisting of dermatologists and neurologists) was consulted shortly prior to the final decision. Some have voiced the patient's perspective. | Responses of the Scientific Advisory Group CNS to the CHMP list of questions on efalizumab 7 Jan 2009. EMA/24463/2009 |
| LINKED DECISIONS<br>12. Consider the consistency of this decision with similar past decisions, and assess whether taking this decision could impact future decisions. | **How this decision might set a precedent or make similar decisions in the future easier or more difficult**:<br>Efalizumab is the first monoclonal antibody ever to be definitively revoked from the market for safety reasons (natalizumab came back with a RMP).<br>The FDA made in US a similar decision to EMA, leading to a US withdrawal of efalizumab from market approximately at the same time as EU and rest of the world.<br><br>Benefit-Risk balance of immunosuppressive monoclonal antibodies with unknown long term effects in non life-threatening diseases with existing alternative treatments may be questionable over time. Development programmes to be adapted to this situation (design, duration, sub population analysis, etc.) | |

## 5.2 BRAT

**Executive Summary**

The PhRMA BRAT 6 step process (ref 3) establishes a structured, transparent framework as a prerequisite for benefit-risk assessment, and can be integrated into the regulatory decision-making process both at the time of drug approval and as new information becomes available. It provides a holistic and progressive qualitative approach to benefit-risk assessment. It is a qualitative approach that has the flexibility to incorporate quantitative elements as needed, can incorporate all relevant aspects of benefit and risk and importantly enriches rather than substitutes for human decision-making

The recommended approaches to applying the Parma BRAT methodology were modified in part to address issues specific to the efalizumab case scenario, but the general principles were followed. Previous work done by the efalizumab team in the creation of the PrOACT-URL case study report was leveraged by KH and DH (efalizumab BRAT sub team).

Since the perspective of the regulator would have been informed by the favourable and unfavourable effects (benefits and risk) data provided via clinical trials and post-marketing surveillance, these effects, as described in the efalizumab regulatory documents ( EPARs, Scientific Discussion, and changes to the Summary of Product Characteristics (SPC)) formed the basis of the initial value tree and the same documents were used to populate the data source table, once it was determined that for a given metric, there was comparative information available for placebo (i.e. background epidemiology rates) . Lack of available data within the time constraints for the case study for placebo, resulted in some unfavourable effects being removed from the initial value tree.

Summary statistics from the data source table were derived with data transformation techniques including a Bayesian mixed effect meta analysis for some of the efficacy measures.

It was decided not to formally weight the outcomes by simulating the regulators perspective on prioritization of the favourable and unfavourable effects. The efalizumab BRAT sub team elected to use the summary data visualizations unadjusted for outcome weightings, to consider the benefit risk balance

The framework, through using the accompanying excel based visualization software delivered two comprehensive visualizations: The key benefit-risk summary table and forest plots, which provided easily interpretable information to inform the benefit risk decision which is taken outside of the framework itself (allowing for the critical element of medical judgment which does not easily lend itself to a quantitative approach.

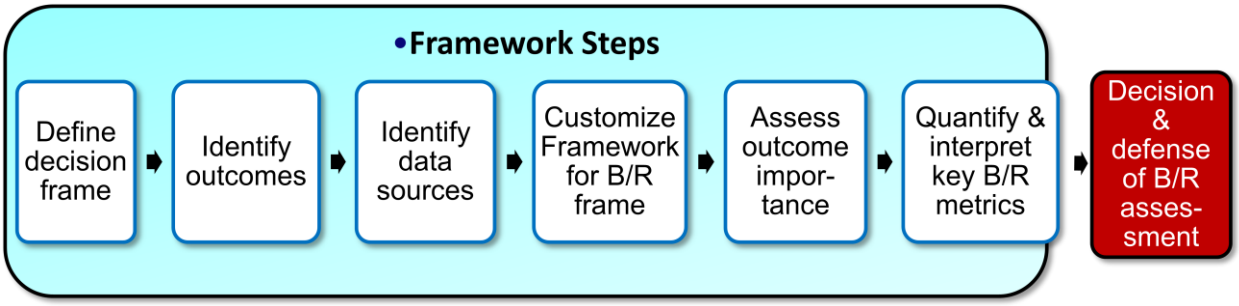
Using the visualizations, and considering the totality of data the efalizumab BRAT sub team decided that the decision really rested on the trade off between the benefits of a PASI 75 effect and the unfavourable effect of PML.

This Implicit "trade off" between PASI 75 and PML provoked the conduction of an additional structured methodology, the benefit risk ratio (BRR).

**Introduction**

The PhRMA BRAT Framework is a set of principles, processes and tools for pharmaceutical B-R decision-making. The principles are guidelines that apply to B-R assessments in general, and the processes and tools serve as a "toolkit" from which decision-makers choose components as appropriate. It is designed for application throughout the life-cycle of a product (from early development to post-approval) and serves both as a tool for sponsors and Health Authorities independently, as well, and almost more importantly as a communication medium and conduit for consensus between them.

The 6 steps of the BRAT framework are shown below:




The actual decision and defense of the B/R decision is taken outside of the 6 step framework. This speaks to the medical judgment that must be integrated at the end of the day to any quantitative activity in order to capture the clinical meaningfulness of the issue at hand. The framework does not issue a decision itself in the form of an integrated B/R summary statistic resulting from the execution of the 6 step process.

The BRAT framework and its modified application to the efalizumab case study is described below.

### 5.2.1 Step 1: Define the decision context

Step one describes the decision context as representing the agreed upon disease state under consideration, the patient population affected, the medicine in question and the appropriate dose and formulation under review, the intended decision and the perspective(s) from which the decision is to be made (patient, regulator or sponsor) and in particular the treatment(s) to which the

medicine in question is being compared (placebo, standard of care, specific competitor medications).

For efalizumab, instead of a separate discussion to determine the decision context, step1 information was essentially drawn from the PrOACT-URL (Step 1 and 3 Problem and Alternatives) output, with additional discussion around comparators, and perspective.

| **Table 1: Decision Context** | |
|---|---|
| Indication | Efalizumab is indicated in the treatment of "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporine, methotrexate and PUVA. The duration of initial therapy is 12 weeks. Therapy may be continued only in patients who responded to treatment (PGA good or better). |
| Drug | Raptiva® (efalizumab) is a recombinant, humanized IgG1 monoclonal antibody that targets CD11a, the α-subunit of leukocyte function associated antigen 1 (LFA-1). Mechanism of action may lead to inhibition of leukocyte migration, similarly to natalizumab.) |
| Formulation/Dose | An initial single dose of 0.7 mg/kg body weight is given followed by weekly injections of 1.0 mg/kg body weight, subcutaneously |
| Comparator | placebo only |
| Population | "high need" adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies |
| Time Frame for Outcomes | 12 weeks for PASI 75 (efficacy/favourable effects), 3 years for PML |
| Perspective | Regulator (at EMA) |

For comparators, consideration was given to methotrexate, cyclosporine and PUVA (available non-biologic therapies at time of initial approval, and also alternative biological therapies (etanercept, infliximab, adalimumab) which became available for the treatment of moderate to severe plaque psoriasis in "high need" patients in the subsequent months after initial approval: however due to time limitations and challenges with data transformation, the comparison was limited to placebo.

The patient perspective would have been a different and important perspective to understand, but requires a formal study employing something akin to the Discrete Choice Experiment , methodology, direct individual patient input or patient representative input. It is also challenging to re create the environment in 2009.

The efalizumab case and retrospective application of the PhRMA BRAT framework highlights the need to formally identify, as part of step one, the chronological reference point, as this would allow for a clearer description of the context when retrofitting decision-making within an historical scenario

### 5.2.2 Step 2: Identify and select benefit and risk outcomes and associated measures

The second step of the PhRMA BRAT framework is to identify and select benefit and risk outcomes and associated measures. The value of any benefit-risk (B-R) assessment depends critically on selecting the appropriate benefit and risk outcomes to be used in the assessment. The PhRMA BRAT guidelines for selecting benefit and risk outcomes include suggestions for specific steps in the process, including (1) identifying the pool of potential outcomes, (2) selecting outcomes to include in the Framework, and documenting decisions made. For prospective applications of the PhRMA BRAT model, these activities are often adjunctive to the identification of primary and secondary endpoints for clinical trials. The approach is primarily qualitative in nature, as often the assessment of which outcomes to be used is done with incomplete or imprecise data.

Besides the efficacy parameters studied during clinical development, other potential benefit outcomes may be identified through literature reviews, regulatory precedents and meetings with clinical experts. The pool of possible outcomes includes all outcomes, whether "known" or "potential", and may include clinical efficacy or effectiveness measures, laboratory measures / biomarkers, certain patient-reported outcome measures / response to treatment, survival, etc. At step 2, not all measures selected will necessarily end up as being important influencers of the B/R decision, but these will be identified in later steps.

All of the inclusion and exclusion criteria must be documented so that there is a transparent and auditable process should decisions require revisiting or defending in the future. For the efalizumab case example, the historical regulatory scenario and decision context specified in Step 1 placed limitations on how step 2 of the PhRMA BRAT framework could be adopted:

Firstly, the perspective of the regulator would have been informed by the favourable and unfavourable effects data provided via clinical trials and post-marketing surveillance. This information was documented in regulatory documents such as EPARs, Scientific Discussion, and changes to the Summary of Product Characteristics (SPC). Therefore the taskforce did not obtain a pool of outcomes from literature reviews, regulatory precedents and meetings with clinical experts as suggested by the PhRMA BRAT method. Instead, regulatory documents were closely examined to specifically address how the regulator would have considered the benefit risk balance with the data which would have been available to them at the time.

Secondly, the PhRMA BRAT framework suggests listing known or potential outcomes relevant to public health, physicians, and patients. This can result in the inclusion of outcomes deemed important by specific stakeholder groups. However, our taskforce refrained from this as we wanted the perspective to be specific to only the regulator and the data they could have accessed at the time of decision-making. This is a traditional perspective, although it is worth noting that explicitly

discussing the outcomes for inclusion between stakeholders can result in a beneficial harmonisation between groups.

Thirdly, the data broadly addressed the outcome of "favourable effects" in terms of efficacy with "unfavourable effects" in terms of safety. The taskforce had previously complied data contained within the regulatory documents into an effects table. However, it became evident that information compiled within effects table placed a primary emphasis on measures rather than outcomes. That is, the data presented to the regulator was often in terms of measures e.g. PASI75, PGA etc., which was then broadly covered with either an umbrella outcome term of "efficacy", or one of "safety". Therefore, we did not perform the suggested task of selecting measures to characterise outcomes, as we had already collected measures of relevance to our scenario

Measures

Table 1 contains a list of measures deemed most likely to influence the benefit-risk balance given the decision context. This list contains all the measures present on the effects table which were considered at the time of decision-making by the regulator; e.g. drug specific safety issues, changes to SPC, reported spontaneous AEs.

**Table 1: Measures most likely to importantly influence the benefit-risk balance**

**Favorable effects:**

- PASI75
- PGA
- OLS
- DLQI
- PASI50

**Unfavourable effects:**

- ADR1 (mild to moderate ADRs as documented in the SPC)
- ADR2 (serious)
- Meningitis aseptic
- Serious infections including pneumonia, sepsis, cellulitis
- Opportunistic infections including fungal infections, tuberculosis, herpes virus infections, EBV, CMV
- Serious thrombocytopenia
- Immune haemolytic anaemia
- Psoriasis severe forms (erythrodermic, pustular)
- Nervous System disorders including Inflammatory polyradiculopathy, Facial Palsy, GBS, Fisher Miller Syndrome
- Interstitial lung diseases including lung infiltration, pulmonary fibrosis
- Serious cases of psoriasis exacerbation or rebound
- Brain infections including Encephalitis and PML

The framework distinguishes between measures which count outcome events and count patients. It is important to note that for each measure within this case study the units varied for the clinical trial and surveillance data presented within regulatory documents. Therefore, there was an extremely

mixed approach to units in the case study, ranging from percent per 100 patient years, to number of cases, to percent. The measures also vary with different time periods and populations.

Composite measures are frequently reported in regulatory documents. Although the framework acknowledges composite measures to be useful when one outcome may not adequately capture the most relevant attribute for a product, it warns that it might introduce difficulty when making important trade-offs in Step 5 of the PhRMA BRAT framework, especially if the outcomes it contains have different effects on patients. Measures are frequently placed together into composite groups within the effects table.

Value tree

It was possible to draw up an initial value tree which contained the benefit and risk measures from Table 1. This tree was further developed in the subsequent steps to only include available, precisely defined, reliable and accurate end-point measures available to the regulator at time of decision-making.

In an unmodified application of the BRAT framework, the tree would exclude outcomes which are considered components of another included outcome, and outcomes which are similar to prevent double counting and overestimation of effects. Our methods deliberately diverged from the method at this point to include all potential outcomes and disregard double counting and overestimation. This was because from a clinical perspective, the larger efalizumab team believed in the importance of measures such as the PASI 75 and PASI 50 for example, (although recognizing that it could be contentious, from the perspective of double counting). This allowed for a full comparison of measures to examine how the visual representation (benefit risk summary table or forest plot) of similar measures may vary.

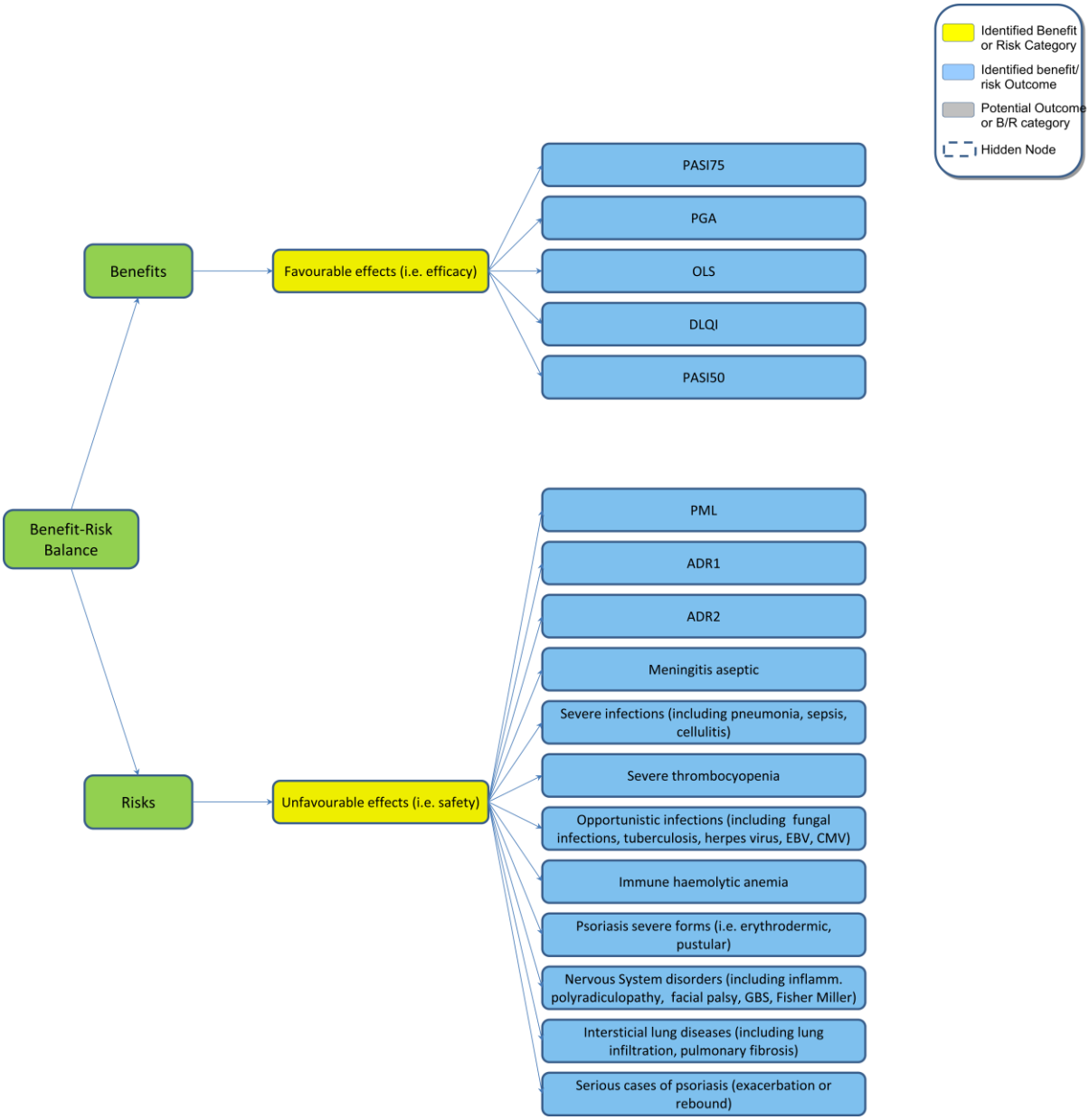



**Fig 1: Initial value tree built using BRAT framework tool (modified)**

### 5.2.3 Step 3: Identify and extract data sources

The third step of the PhRMA BRAT framework describes the identification and selection of data sources, in addition to organising them and extracting the relevant data. For the case study of efalizumab, any document which would have been available to the decision-maker, i.e. the regulator, was included if it was publically available at the time of decision-making, or could be publically provided upon request. Regulatory documents containing favourable and unfavourable information from clinical trials and post-marketing surveillance were identified, and documented. The relevant data were then extracted.

Inclusion of measures

Rationale for inclusion or exclusion of data was documented (Table 2). Specifically, measures were only included if they had sufficient information to complete the required data source table fields in Step 4, e.g. background epidemiology of placebo known.

**Table 2: Measures and inclusion**

| Measure | Source | Inclusion | Rationale |
|---|---|---|---|
| PASI75 | Clinical trials | Yes | Complete data |
| PGA | Clinical trials | Yes | Complete data |
| OLS | Clinical trials | Yes | Complete data |
| DLQI | Clinical trials | No | Average and standard deviation missing |
| PASI 50 | Clinical trials | Yes | Complete data |
| ADR1 | ISS | Yes | Complete data |
| ADR2 | ISS | No | Percentage of events in placebo group not given; percentage of events for efalizumab not precise (range given) |
| Meningitis aseptic | PSUR10 | No | Background epidemiology not known |
| Serious infections including pneumonia, sepsis, cellulitis | ISS | Yes | Complete data |
| Opportunistic infections including fungal infections, tuberculosis, herpes virus infections, EBV, CMV | PSUR10 | No | RMP only states background epidemiology of tuberculosis; background epidemiology of other conditions not known |
| Serious thrombocytopenia | PSUR10 | No | Background epidemiology not known |
| Immune haemolytic anaemia | PSUR10 | No | Background epidemiology not known |
| Psoriasis severe forms (i.e. erythrodermic, pustular) | ISS | Yes | Complete data |

| Table 2: Measures and inclusion | | | |
|---|---|---|---|
| **Measure** | **Source** | **Inclusion** | **Rationale** |
| Nervous System disorders including Inflammatory polyradiculopathy, Facial Palsy, GBS, Fisher Miller syndrome | PSUR10 | No | Background epidemiology not known |
| Interstitial lung diseases including lung infiltration, pulmonary fibrosis | PSUR10 | No | Background epidemiology not known |
| Serious cases of psoriasis exacerbation or rebound | PSUR10 | No | Background epidemiology not known |
| Brain infections including Encephalitis and PML | PSUR10 | Yes | Complete data |

Data source table

The PhRMA BRAT guidelines state that a data source table should be completed, which includes all study and publication details. Such a table would ideally store sufficient information and relevant metrics to enable a full and articulate discussion of benefit-risk while providing complete transparency into the origin and format of the source data, The visualizations software which was provided to users of the BRAT framework is an excel based application that does not have the capacity to store information on the data sources, and instead contains a spreadsheet to store details of the measures (Table 3).

**Table 3: Data table**

| Outcome name | Treatment 1 rate point estimate | Treatment 1 rate lower CI | Treatment 1 rate upper CI | Treatment 1 number of patients | Treatment 1 number of events | Duration treatment 1 |
|---|---|---|---|---|---|---|
| PASI75 | 0.28 | 0.18 | 0.41 | 1742 | 485 | 12 |
| PGA | 0.3 | 0.18 | 0.46 | 1742 | 531 | 12 |
| OLS | 0.29 | 0.18 | 0.44 | 1742 | 508 | 12 |
| PML | $8.51\ 10^{-5}$ | $1.72\ 10^{-6}$ | $1.69\ 10^{-4}$ | 47000 | 4 | PMS |
| ADR1 | 0.41 | 0.39 | 0.43 | 1742 | 714 | 12 |
| Psoriasis severe forms | 0.03 | 0.02 | 0.04 | 1742 | 56 | 12 |
| **Outcome name** | **Treatment 2 rate point estimate** | **Treatment 2 rate lower CI** | **Treatment 2 rate upper CI** | **Treatment 2 number of patients** | **Treatment 2 number of events** | **Duration treatment 2** |
| PASI75 | 0.04 | 0.02 | 0.06 | 979 | 36 | 12 |
| PGA | 0.05 | 0.03 | 0.09 | 979 | 51 | 12 |
| OLS | 0.04 | 0.02 | 0.06 | 979 | 36 | 12 |
| PML | $4.40\ 10^{-6}$ | $3.10\ 10^{-6}$ | $5.70\ 10^{-6}$ | 10000000 | 44 | PMS |
| ADR1 | 0.24 | 0.21 | 0.27 | 979 | 235 | 12 |
| Psoriasis severe forms | 0.01 | 0.01 | 0.02 | 979 | 14 | 12 |
| **Outcome name** | **Risk difference point estimate** | **Risk difference lower CI** | **Risk difference upper CI** | **Relative risk point estimate** | **Relative risk lower CI** | **Relative risk upper CI** |
| PASI75 | 0.24 | 0.15 | 0.36 | 7.82 | 5 | 12.38 |
| PGA | 0.25 | 0.14 | 0.4 | 5.78 | 3.6 | 9.34 |
| OLS | 0.25 | 0.15 | 0.39 | 7.81 | 4.73 | 13.27 |
| PML | $8.07\ 10^{-5}$ | $-2.70\ 10^{-6}$ | $1.64\ 10^{-4}$ | 19.34 | 6.95 | 53.83 |
| ADR1 | 0.17 | 0.13 | 0.21 | 1.71 | 1.51 | 1.93 |
| Psoriasis severe forms | 0.02 | 0.01 | 0.03 | 2.25 | 1.26 | 4.02 |

Creation of the summary tables and graphs, by definition, requires summarizing the benefit / risk outcome (favourable/unfavourable effects) data across multiple studies. There is a wide range of units of measure for individual study results that may be entered into the framework including absolute risk difference, relative risk, odds ratio, incidence, adjusted relative risk, time to onset, etc.

For efalizumab, with the data available for each measure as described in the pertinent regulatory documents, 95% confidence intervals, point estimates, risk differences, and relative risks were calculated using the formulae listed in Table 4. Additional data transformation included a Bayesian mixed effects meta analysis performed for PASI75, PGA and OLS.

The challenge of adequately representing a rare event in the post marketing setting in a manner that translates to the other effect measures, is seen with PML. As we would expect with such small numbers, the relative risk metric is significant but with some uncertainty as evidenced by the wide

confidence interval. This metric also bears the statistical assumption that there is a consistent rate of PML across time that may not be true for PML in this patient population.

**Table 4: Definitions and equations used within the data source table**

| Column | Description | Formula |
|---|---|---|
| Outcome | Outcome of interest | |
| Study | Name/code of study | |
| Treatment 1 rate point estimate ($TR_R$) | Probability of having an event in the efalizumab arm of the trial | $ev_{arm}/n_{arm}$ |
| Treatment 1 rate lower/upper CI | Gives the upper and lower confidence intervals of the treatment rate point estimate for efalizumab arm | $TR_R \pm 1.96\sqrt{\frac{TR_R(1-TR_R)}{n_R}}$ |
| Treatment 1 number of patients ($n_R$) | Number of patient in the efalizumab arm | |
| Treatment 1 number of events ($ev_R$) | Number of events of specified outcome | |
| Treatment 2 rate point estimate ($TR_P$) | Probability of having an event in the Placebo arm of the trial | |
| Treatment 1 rate lower/upper CI | Gives the upper and lower confidence intervals of the treatment rate point estimate for Placebo arm | $TR_P \pm 1.96\sqrt{\frac{TR_P(1-TR_P)}{n_P}}$ |
| Treatment 1 number of patients ($n_P$) | Number of patient in the Placebo arm | |
| Treatment 1 number of events (ev) | Number of events of specified outcome | |
| Risk difference point estimate (RDiff) | Difference in risk of having specified event between efalizumab arm and Placebo arm | $TR_R - TR_P$ |
| Risk difference lower/upper CI | Gives the upper and lower confidence interval of the risk difference point estimate | $RDiff \pm 1.96\sqrt{\frac{ev_R(n_R-ev_R)}{n_R^3}+\frac{ev_P(n_P-ev_P)}{n_P^3}}$ |
| Relative risk point estimate ($RR$) | Is the relative risk of developing specified outcome in the efalizumab arm when compared to the placebo Arm | $RR = \frac{(ev_R/n_R)}{(ev_P/n_P)}$ |
| Relative risk lower/upper CI | Gives the upper and lower confidence interval of the relative risk point estimate | $e^{\ln RR \pm 1.96\sqrt{\left(\frac{1}{ev_R}-\frac{1}{n_R}\right)+\left(\frac{1}{ev_P}-\frac{1}{n_P}\right)}}$ |

### 5.2.4 Step 4: Customise framework

Step four customises the framework. The initial value tree created in step two is modified to account for clinical expertise and the data reviewed in step three. Outcomes considered irrelevant to the benefit- risk assessment or stakeholder groups are either refined to obtain relevance or removed.

Application to case study

For our case study, we were constrained by the quality of data sources. Measures with incomplete data (e.g. no details on background epidemiology) were removed. Table 5 lists the final outcome measures used.

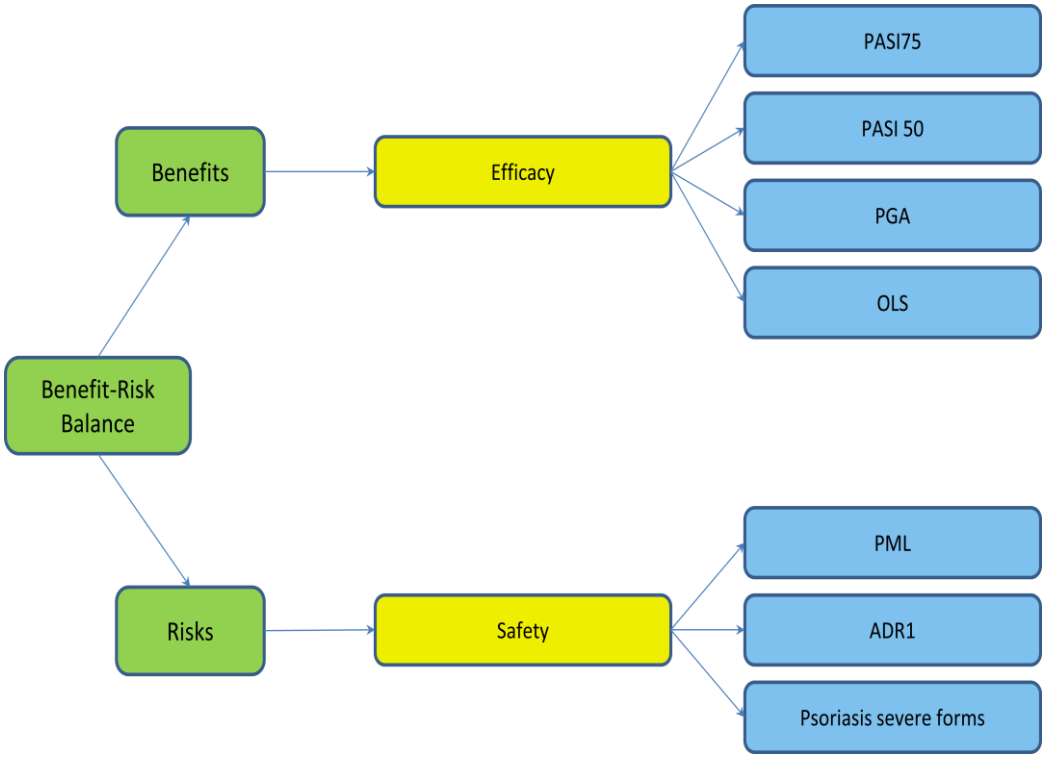
| **Table 5: Final list of outcome measures** | |
|---|---|
| PASI75 | Proportion of patients who achieve a 75% reduction in PASI scores. The PASI score is derived by evaluating erythema, scaling and thickness and then weighting the coverage according to the area covered, i.e. head, trunk, upper extremities and lower extremities. The scores can range from 0 (least severe) to 72 (most severe). |
| PASI50 | Proportion of patients who achieve a 50% reduction in PASI scores after two weeks. See PASI75 scoring of the PASI. |
| PGA | Static PGA is a measure of the psoriatic lesions taken at a single time point. The scores can range from 7 (least severe) to 1 (most severe). |
| OLS | The OLS is a global rating of psoriasis severity according to plaque elevation, scaling, and erythema at a given time point. The scores can range from 0 (least severe) to 5 (most severe). |
| PML | Progressive multifocal leukoencephalopathy. A demyelinating disease caused by reactivation of the John Cunningham virus. |
| ADR1 | Mild to moderate dose related acute flu-like symptoms. |
| Psoriasis severe forms | E.g. erythrodermic, pustular |

As stated in the previous step, there are differences in the design, and outcome measures between clinical trials and post-marketing surveillance. For example, two of the clinical trials pooled the data for efalizumab 1mg/kg/wk, and efalizumab 2mg/kg/wk when calculating unfavourable effects for examples such as ADR1 and psoriasis severe forms. It is assumed the effect will be small and result in a minor overestimation of adverse events. Additionally, the follow up time for clinical trials was

set to twelve weeks, whereas the follow up time for post-marketing surveillance was cumulative and lasted for 47,000 patient-years.

Tuning was made accordingly with the data available and is displayed below (Figure 2).

**Figure 2: Modified value tree**



### 5.2.5 Step 5: Assess outcome importance

In this step, outcomes are assessed for their importance to the decision-maker and other stakeholders, and the subsequent rankings and weightings are applied to the decision tree. When outcomes are differentially weighed relative to one another, it allows for a transparent discussion on priorities between different stakeholder groups and can provide the basis for a sensitivity analysis over the different perspectives. It is important to note that the PhRMA BRAT framework does not advocate a specific method to weigh the preferences of outcomes in the value tree, or require the use of weights at all.

For the purposes of the efalizumab case study, and in the context of PhRMA BRAT framework, it was decided not to formally weight the outcomes by simulating the regulators perspective on prioritization. The efalizumab BRAT sub team elected to use the summary data visualizations unadjusted for outcome weightings, to consider the benefit risk balance (having selected the outcomes most likely to influence the benefit risk determination, outcomes which were in turn supported by appropriate data)

### 5.2.6 Step 6: Display and interpret key benefit-risk metrics

7 places the source data into a key benefit–risk summary table which summarises the key information in source data required to quantify outcomes in the value tree. The table aids interpretation of benefits and risks; the use of such a framework can increase the transparency, predictability and consistency with which benefit-risk assessments are conducted.

**Table 7: Key benefit-risk summary table**

| | | Outcome | RAPTIVA Risk / 1000 pts | Placebo Risk / 1000 pts | Risk Difference (95% CI)/ 1000 pts | | Relative Risk (95% CI) | |
|---|---|---|---|---|---|---|---|---|
| Benefits | Efficacy | PASI75 | 280 | 36 | 244 | (151, 362) | 7.819 | (4.999, 12.380) |
| | | PASI 50 | 567 | 200 | 360 | (303, 431) | 2.800 | (2.210, 3.650) |
| | | PGA | 305 | 52 | 251 | (141, 396) | 5.778 | (3.602, 9.337) |
| | | OLS | 292 | 37 | 254 | (145, 392) | 7.813 | (4.731, 13.270) |
| Risks | Safety | PML | 0 | 0 | 0 | (0, 0) | 18.400 | (5.400, 45.960) |
| | | ADR1 | 410 | 240 | 170 | (130, 210) | 1.710 | (1.510, 1.940) |
| | | Psoriasis severe forms | 33 | 15 | 17 | (6, 29) | 2.170 | (1.270, 3.970) |

The PhRMA BRAT framework delivers two comprehensive visualizations: the key benefit-risk summary table and forest plot >; these provide easily interpretable information to stakeholder groups-- such as patients and healthcare professionals, enabling them to make informed decisions based on their own preferences. Use of colour-coding throughout the graphs and tables is effective in differentiating among types of data (e.g., benefits vs. risks), and for highlighting certain results.

Tables and graphs that seem to communicate effectively are those that display a fairly limited number of data columns, e.g., 5-6 depending on the audience, interpreting odds ratios and 95% confidence intervals can also be challenging.

Forest plot

The forest plot records all of the measures on a standardised scale, allowing for the evaluation of each measure relative to other measures. Below are two forest plots comparing efalizumab with placebo, firstly with risk differences for the selected benefit risk outcome measures and secondly with relative risk for the same outcome measures.

**Figure 3: Forest plot: Risk difference for key favorable and unfavorable effects (**efalizumab **compared to placebo)**

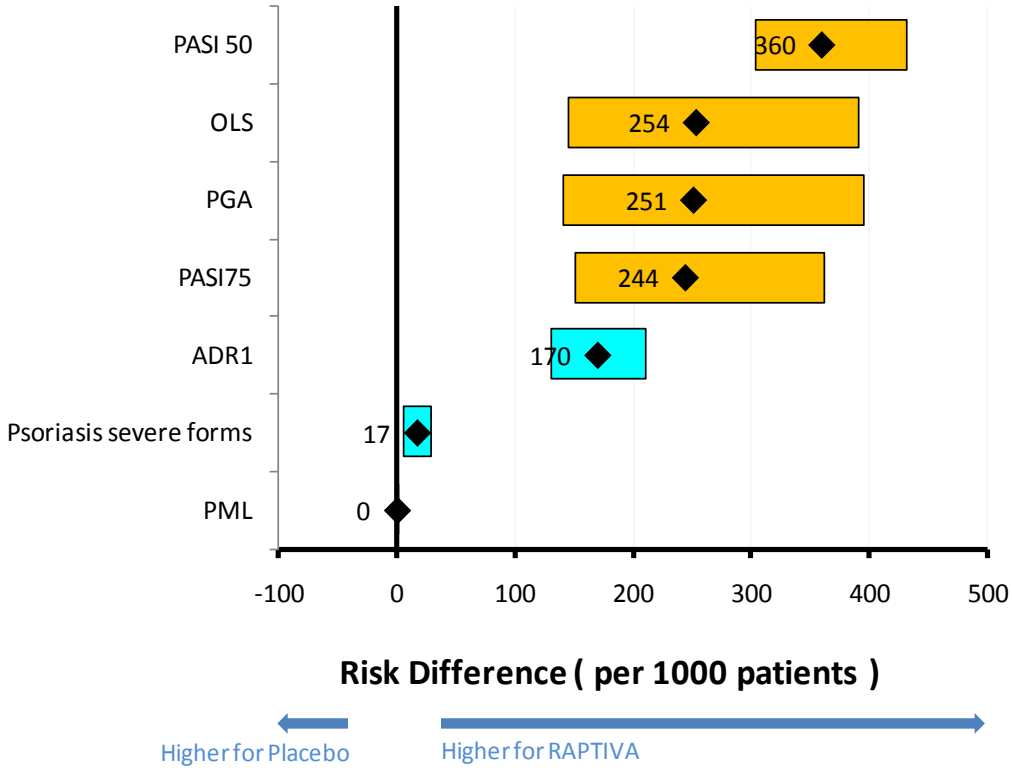
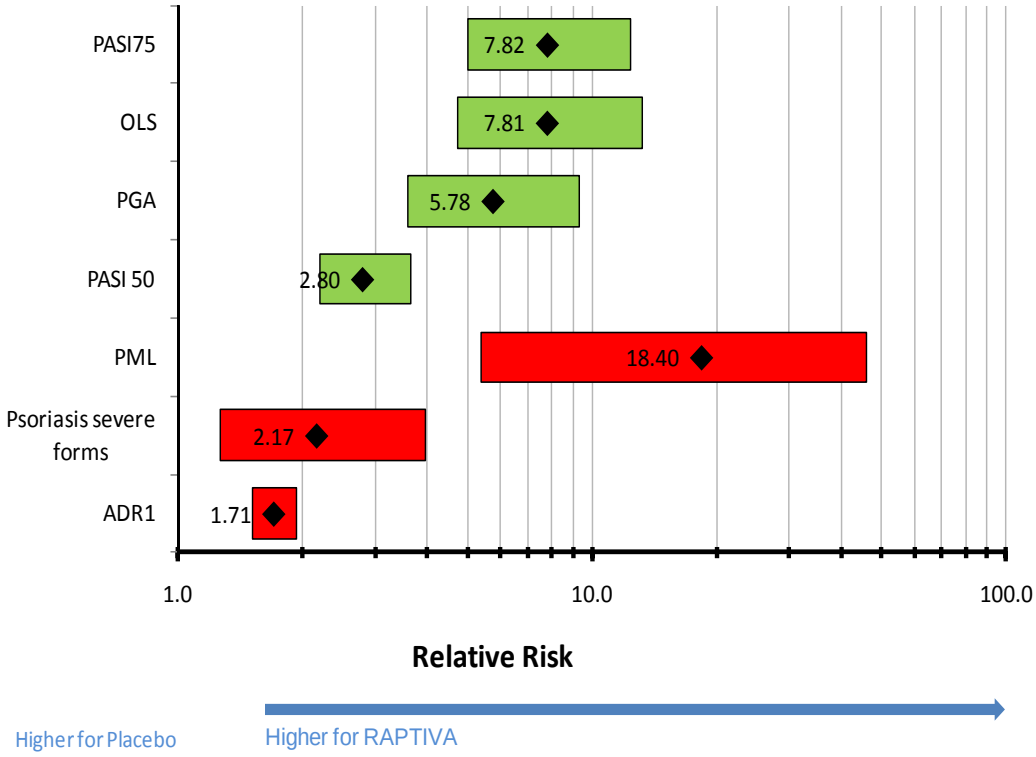


**Figure 4: Forest plot: Relative Risk for key favorable and unfavorable effects (**efalizumab **compared to placebo)**




## Benefit/risk determination

The summary information, (both the key benefit risk summary table and forest plots) will lead individuals to draw different conclusions from the same data. In January 2009, when the Regulators at the EMA were considering if there were any risk minimisation measures which could be rapidly implemented, to maintain a positive B-R balance for efalizumab and this avoid Market Authorisation suspension, a structured approach to a benefit/risk analysis applying the PhRMA BRAT framework on the data available to them, would have yielded the above visualizations.

Looking at the modest efficacy in terms of the differences in risk or relative risk between efalizumab and placebo for the key outcomes, and considering this in the context of the risk of PML and other unfavourable effects such as rebound psoriasis, the regulators would be thinking about the population of adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies, knowing that there are now alternative biological therapies available to treat the same condition, therapies which have a better defined safety profile by virtue of the incremental exposure resulting from additional indications.

The efalizumab BRAT sub team (KH and DH), when considering the totality of data (including the unfavourable effects such as opportunistic infections, other serious infections, serious thrombocytopenia, and some nervous system disorders which were removed at step 4 of the PhRMA BRAT framework process because of lack of data for placebo), determined that in the end the decision really rested on the trade off between the benefits of a PASI 75 effect and the unfavourable effect of PML. Instinctively we would have voted for market authorization suspension in the absence of robust mitigation measures for PML.

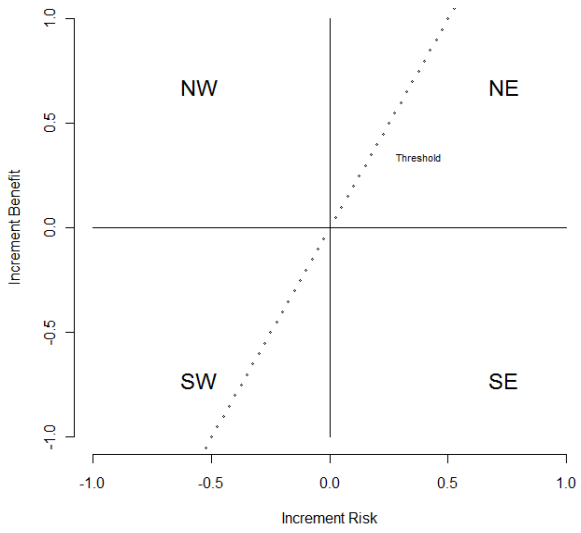
This Implicit “trade off” between PASI 75 and PML provoked the conduction of an additional structured methodology, the benefit risk ratio (BRR).

## 5.2.7. Conduct of efalizumab Benefit Risk analysis using BRR (Benefit-Risk ratio); Benefit: PASI 75; Risk: PML incidence.

### *5.2.7.1: BRR method illustration:*

BRR considers the increment of benefit (testing drug over its comparator or placebo) against the increment of risk. The risk increment $\Delta R$, and benefit increment $\Delta B$, will be taken as x-coordinate and y-coordinate into the following x-y plane

**Figure 5. Benefit-Risk plane and threshold line**




If $(\Delta R, \Delta B)$ falls in 'NW' quadrant, the drug under consideration is dominant over its comparator. If $(\Delta R, \Delta B)$ falls in 'SE' quadrant, the comparator is dominant over the drug under consideration. If $(\Delta R, \Delta B)$ falls in 'NE' quadrant, which is the most frequently observed, the decision depends on a pre-fixed choice of 'threshold' $\mu$. If $\left(\frac{\Delta B}{\Delta R}\right) > \mu$, which means that when point $(\Delta R, \Delta B)$ is over the threshold line, the drug under consideration is considered as positive, otherwise the drug under consideration is considered as negative.

### *5.2.7.2: BRR analysis for efalizumab case study:*

In the section, BRR method will be applied to the efalizumab case study. Since BRR deals with two criteria only, we chose PASI75 as the primary benefit criterion and PML as the primary risk criterion in this analysis. All secondary benefits and risks are excluded from the analysis.

#### 5.2.7.2.1 Benefit increments (efalizumab - Placebo)

Pooling five clinical trials in a Bayesian meta-analysis gives the increment of PASI75 benefit

$$\Delta B = 0.2450 \ \ (95\%CI: 0.2224, 0.2676)$$

Beware that the PASI75 effect is measured 12 weeks post treatment. In this analysis we assume this effect does not change over time.

#### 5.2.7.2.2 PML risk increments (efalizumab - Placebo) under two exposure estimations

The increment of PML risk is however time dependent. The data (PML document) shows clearly an increasing trend of PML risk over time. The patient exposure is hard to estimate precisely, so a low estimation and a high estimation are given in PML risk table.

Estimated number of patients and PML incidence by exposure

| Exposure to efalizumab | Estimated number of patients<br>Low - High | PML cases | Incidence (per 1'000 pts) for high and low exposure estimates<br>Range (95%CI) | |
|---|---|---|---|---|
| On drug | 40'000 – 60'000 | 0 at <1 year | 0<br>(0-0.06) | 0<br>(0-0.09) |
| On > 1 year | 15'454 – 23'180 | 0 at 1-2 years | 0<br>(0-0.16) | 0<br>(0-0.24) |
| On > 2 years | 5'944 – 8'914 | 0 at 2-3 years | 0<br>(0-0.41) | 0<br>(0-0.62) |
| On > 3 years | 2'236 – 3'832 | 3 at 3-4 years | 0.78<br>(0.16-2.29) | 1.34<br>(0.28-3.92) |
| On > 4 years | 856 – 1'738 | 1 at >4 years | 0.58<br>(0.01-3.20) | 1.17<br>(0.03-6.49) |

With low exposure estimation, the PML risk after year 3, from year 3 to year 4, is 1.34 over 1000 (pts). The increment of PML risk from placebo is

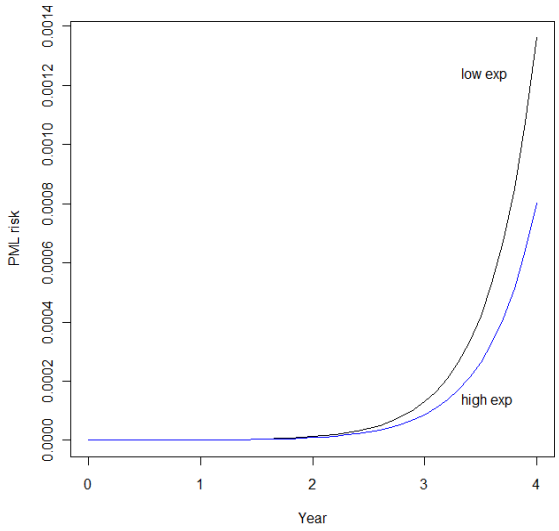
$$\Delta R_{le} = 1.34 \times 10^{-3} \ \ (95\%CI: 0.28 \times 10^{-3}, 3.24 \times 10^{-3}).$$

With high exposure estimation, the PML risk after year 3, from year 3 to year 4, is 0.78 over 1000 (pts). The increment of PML risk from placebo is

$$\Delta R_{he} = 0.78 \times 10^{-3} \ \ (95\%CI: 0.16 \times 10^{-3}, 1.89 \times 10^{-3}).$$

For the PML risk before year 3, there is no observation of PML, so the estimation directly from data is 0. This will lead to a negative risk increment (although very tiny) since we have a positive estimation of background risk. In this analysis, we fit the data (number of PML cases and exposure each year) with an exponential risk increase model, which is more likely than fixed risk model and linear risk increase model in view of the data. The fitted curve of low exposure risk and curve of high exposure risk are shown bellow.

**Figure 6: PML risk increase with time: low exposure risk (black) and high exposure risk (blue)**




Based on this fitting, the PML risk increment in year 2 to year 3 for low exposure scenario is

$$\Delta R_{le} = 0.08 \times 10^{-3} \ \ (95\%CI: 0.02 \times 10^{-3}, 0.20 \times 10^{-3})$$

The PML risk increment in year 2 to year 3 for high exposure scenario is

$$\Delta R_{he} = 0.06 \times 10^{-3} \ \ (95\%CI: 0.01 \times 10^{-3}, 0.13 \times 10^{-3}).$$

#### 5.2.7.2.3 Ratio of benefit increment and risk increment

The ratio of benefit increment and risk increment for year 3 - 4 in low and high exposure cases are

$$\frac{\Delta B}{\Delta R_{le}} = \frac{0.245}{1.34 \times 10^{-3}} = 182.8 \qquad \frac{\Delta B}{\Delta R_{he}} = \frac{0.245}{0.78 \times 10^{-3}} = 314.1$$

With the PML risk fitting in Figure 2, the ratio for year 2 - 3 in low and high exposure cases are

$$\frac{\Delta B}{\Delta R_{le}} = \frac{0.245}{0.08 \times 10^{-3}} = 3062.5 \qquad \frac{\Delta B}{\Delta R_{he}} = \frac{0.245}{0.06 \times 10^{-3}} = 4083.3$$

The ratio for year 1 - 2 and ratio for year 0 - 1 are much higher.

#### 5.2.7.2.4 Threshold determination

The threshold in this case study can be determined by considering question: what is benefit increment in order to tolerate 0.1% PML risk increase (or say, 1 more PML per 1000 patients). Consider the answer of different choices: 10%, 20% and 60% (i.e., Out of 1000 patients, 100, 200 or 600 more patients benefit from efalizumab than placebo can compensate the risk of 1 more PML from efalizumab than placebo). The three thresholds are then

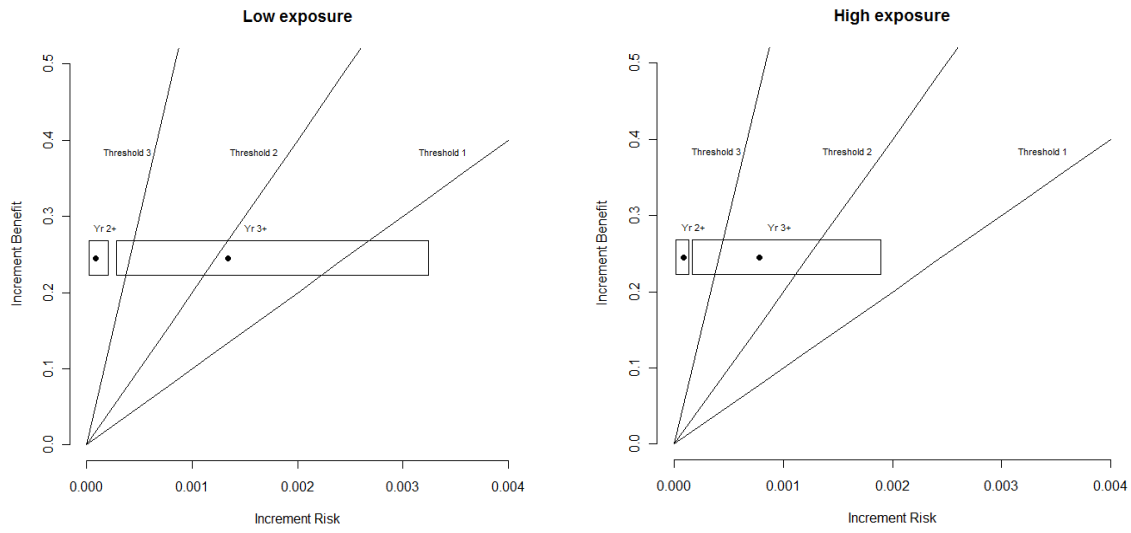
$$\mu_1 = \frac{100 \times 10^{-3}}{1 \times 10^{-3}} = 100 \quad \mu_2 = \frac{200 \times 10^{-3}}{1 \times 10^{-3}} = 200 \quad \mu_3 = \frac{600 \times 10^{-3}}{1 \times 10^{-3}} = 600.$$

Around these three thresholds, the first one is risky, the second may be fine but is still a little bit risky, and the third is a safe choice (maybe a little conservative).

#### 5.2.7.2.5 BRR decision

The decision is clearly dependent on which threshold a decision maker is going to use. The position of $(\Delta R, \Delta B)$ for year 3 - 4 and year 2 - 3 and threshold lines are plotted in following figure.

**Figure 7:** efalizumab **study** $(\Delta R, \Delta B)$ **for year 3 - 4 and year 2 - 3**




The boxes in this figure are confidence range of position$(\Delta R, \Delta B)$. Overall, year 2 - 3 show positive profile for all thresholds. Year 3 - 4 is however complicated. $(\Delta R, \Delta B)$ of year 3 - 4 is over the middle threshold for high exposure estimation, but is under the middle threshold for low exposure estimation. Looking at the figure, an aggressive decision maker may approve the drug, but a conservative decision maker may reject the drug.

## 5.3 MCDA

### 5.3.1 Executive Summary

During a decision conference (ref 4) at the EMA on 1 December 2011, seven members of the efalizumab Case Study Team of the PROTECT project developed a decision-theory-based model for evaluating the benefit-risk balance of efalizumab, a drug for the treatment of adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to other systemic therapies, compared to a placebo. The decision conference took the view of regulators in early 2009, when they were assessing the benefit-risk balance in light of new information received post-authorisation. This report summarises the process and results of the decision conference.

The group considered five favourable effects and ten unfavourable effects, the latter representing five effects from the clinical trials, and five from post-marketing observational data (p. 52, Effects Tree, Figure 1). Each criterion was carefully defined to enable meaningful evaluations of the drugs (pp. 54-55, Effects Table, Table 1). Measurement scales used in the clinical studies were identified for all the criteria.

Pooled data from phase III studies provided measures on the five favourable effects and the five unfavourable effects criteria observed in the clinical trials. Data for the five observational criteria were taken from the Merck Serono PSUR 10 document. Measures for each criterion were converted to preference values on 0-100 scales that were defined as encompassing the range of data, plus possible uncertainties, for each criterion (p. 54-55, Table 1 gives the ranges; page 57, Figure 3 provides an example). All conversions of measures to preferences employed direct or inverse linear transformations, except for PML, for which an inverse convex value function was judged by participants' to capture the clinical relevance of this effect (p. 56, Figure 2). All input scores are show in the Effects Table, while their associated preference values are shown in Appendix B in the "Supplement 2 to Wave 1 case study report Efalizumab" (http://www.imi-protect.eu/benefitsRep.shtml).

The group also assessed relative weights for all the criteria. These weights equate the units of preference value across all the criteria. The method of swing weighting, which requires comparative judgments about the ranges of effects and clinical judgements about how much they matter relative to each other, made it possible to assign meaningful relative weights to all scales (p. 61, Figure 7). These weights reflect both the range from the least to most preferred effects on each scale, a matter of fact, and how much those effects differences matter, a consideration of clinical relevance that takes the context for decision making into account. The model's separation of facts from judgements ensures that swing-weights are scale constants, whereas the more commonly-asked

question “how important is this effect compared to that one”, does not yield meaningful scale constants.

Weighted averages of the scores, calculated by a computer and projected on-the-spot for the group as the model was constructed, provided a single, overall score for each treatment, with efalizumab scoring 51 (out of a possible 100—which would indicate maximum scores on all the favourable effects and no unfavourable effects), and the placebo 31, showing that the drug is overall most preferred.

Those scores are broken down into their favourable and unfavourable effect contributions (p. 63, Figure 9) or by the contributions of the individual criteria (p. 64, Figure 10). Comparisons of the drug with the placebo showed that the main advantages of the drug are the PGA and the PASI75, while the main disadvantage is its potential for PML (p.65, Figure 12). It is this latter display that is perhaps the most useful to regulators and assessors as it shows the differences between drug and placebo based on both the measured data, whatever its form (percentages, scores, change scores, etc.) and the clinical relevance of the data.

Sensitivity analyses showed that the model is very robust to very substantial changes in individual weights for all criteria except PML. The key trade-off is between the 0-60% range on the PASI75 scale and the 0-5 range on PML, which was initially judged to be in the ratio of 2 to 1 (p. 60, Figure 6). Changing that ratio to be about equal, i.e., 60% of patients experiencing a 75% reduction in baseline PASI judged to be as clinically desirable as 5 cases of PML is undesirable, causes the overall benefits to be just balanced by the overall risks. Further increasing the weight on the PML scale causes the risks to exceed the benefits.

Modelling efalizumab at this point in time, two years after the drug was withdrawn, proved to be difficult because the judgements made in 2009 by the assessors and regulators are not recoverable. It is not even possible to know precisely what data led regulators to their decision, for none of the public documents, from 2004 onward, are clear about which criteria the assessors considered relevant to the benefit-risk balance, and which were not. So, though it was possible to model efalizumab retrospectively, the model developed here may well be an incomplete representation of all the explicit and implicit considerations assessors brought to bear at the time the assessment reports were written.

### 5.3.2 Efalizumab Benefit Risk Appraisal

This report documents the process and results of a decision conference (a group modelling process described in Appendix A in the "Supplement 2 to Wave 1 case study report Efalizumab" (http://www.imi-protect.eu/benefitsRep.shtml)) on 1 December 2011 whose purpose was to create and explore a model of the benefit-risk balance for the drug efalizumab. The drug received marketing authorisation on 20 September 2004 for the treatment of adult patients with moderate to severe chronic plaque psoriasis who have failed to respond to other systemic therapies. By January 2009 the margin of benefits over risks had narrowed since approval, so the European Commission requested the CHMP to assess the concerns and its impact on the benefit/risk balance for efalizumab, to give its opinion on measures necessary to ensure the safe and effective use of efalizumab, and on whether the marketing authorisation for this product should be maintained, varied, suspended or withdrawn. The Marketing Authorisation Holder (MAH) did not wish to conduct further clinical trials, as the CHMP had required lifting the suspension recommended in February, so in June the European Commission withdrew the marketing authorisation for efalizumab.

This decision conference took the view of regulators in early 2009, when they were assessing the benefit-risk balance in light of the new information received post-authorisation. Two sources of data contributed to the benefit-risk model: the original 2004 EPAR and the PSUR 10 document provided by Merck Serono[1]. This report summarises the structure of the model developed at the decision conference and the results.

### 5.3.3 Model Structure

After a brief overview by Larry Phillips of the nature and purpose of a decision conference, he reminded participants of the primary task for the day: to develop a benefit-risk model of efalizumab, assuming a regulator's perspective in early 2009. AM and KH had assembled the relevant data from the EPAR and PSUR into an extended Effects Table, which summarised the benefit and risk criteria as favourable and unfavourable effects, with their definitions, the relevant patient population from which the data were drawn, the measurement scales associated with the criteria, the units of measurement and the data. The Effects Table was created during the application of the PrOACT-URL framework to the modelling of efalizumab. This pre-work expedited the work of the group in building a model.

#### *5.3.3.1 The Options*

The group recognised that data were available only for two options:

[1] PSUR 10 was the last Periodic Safety Update Report submitted to EMA in November 2008 before Market Authorisation suspension in February 2009.

1. efalizumab in 2009 (pre and post-marketing data)
2. Placebo in 2004 (premarketing data)

No data were available for an option discussed at the time both by Regulators and Company, resulting in a limitation of treatment to 2 years.

#### *5.3.3.2 The Criteria*

Five favourable effects and ten unfavourable effects characterise the final model. The clinical trials conducted prior to approval provided data for the five favourable effects and for five of the unfavourable effects, while the Merck Serono PSUR 10 document provided data for the other five unfavourable effects. Although the available documentation reports many effects, the group chose to model only those effects that might affect the benefit-risk balance; thus, many unfavourable effects were not included in the model. The Effects Tree, Figure 1, shows favourable and unfavourable effects at the nodes, and criteria against which the drugs are evaluated at the extreme right.

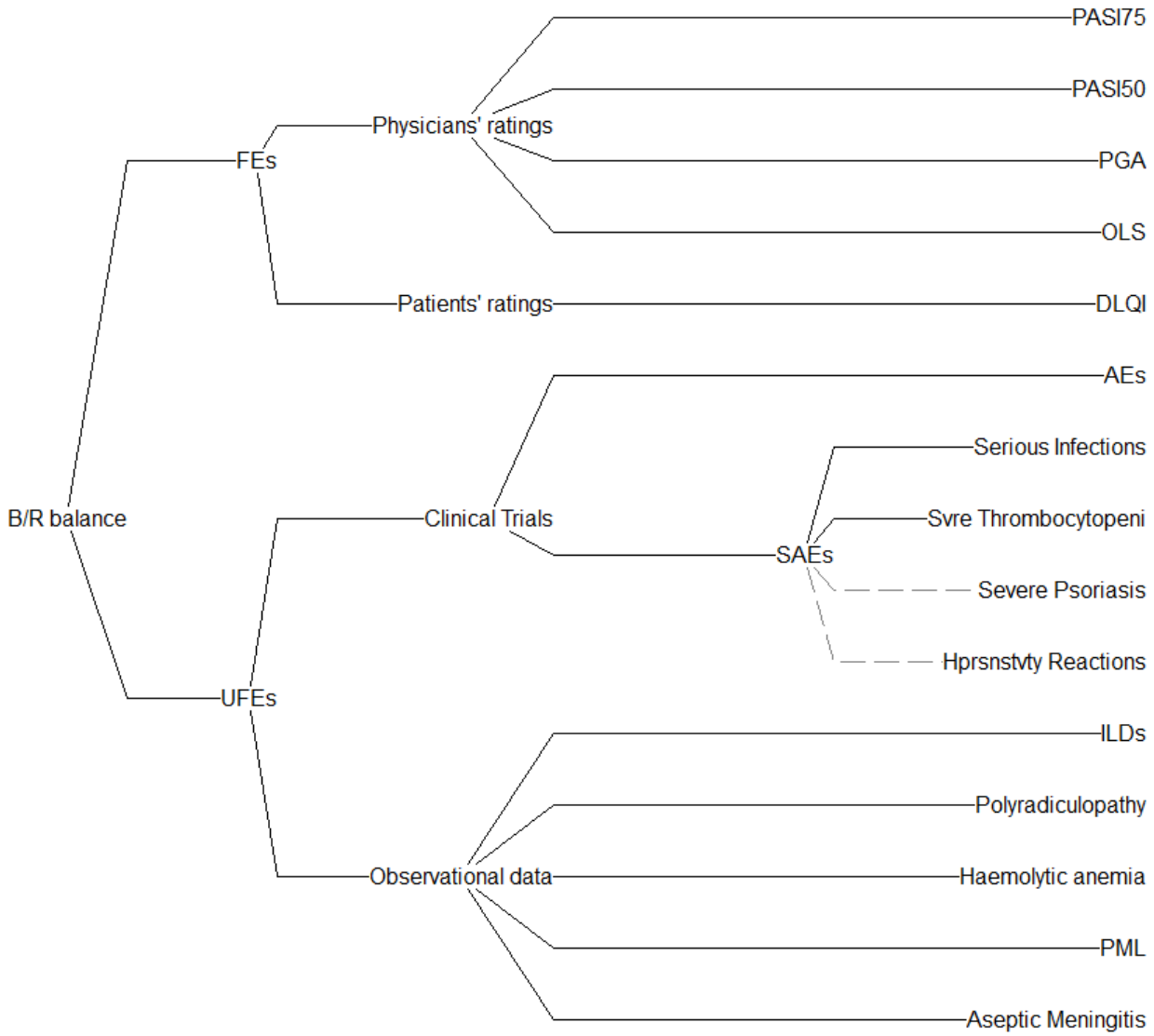



**Figure 1: The evaluation criteria organised by Favourable Effects (FE) and Unfavourable Effects (UFE). The weights assigned to Severe Psoriasis and Hypersensitivity Reactions were so small that their cumulative weights are effectively zero, indicated by the dashed lines.**

An analysis of the data after the decision conference showed that although Serious Infections and Severe Thrombocytopenia were reported in the PSUR, they were less prevalent than in the clinical trials, where the model showed they had no effect on the benefit-risk balance, so they were not included as relevant criteria for the Observational Data.

Definitions of the criteria are given in Effects Table, Table 1. The table shows the short name given in Figure 1, the description of the effect, which in some cases is further explained in the footnotes, fixed upper and lower values that define a plausible range for the data, the units of measurement, and, finally, the data for efalizumab and the placebo. Data from more than one clinical trial were pooled to give the values shown in the Effects Table.

**Table 6: Effects Table for** efalizumab**.**

| | Name | Description | Fixed Upper | Fixed Lower | Units | Raptiva | Placebo |
|---|---|---|---|---|---|---|---|
| Favourable Effects | PASI75 | Percentage of patients achieving 75% reduction in baseline PASI[1] at week 12. | 60.0 | 0.0 | % | 29.5 | 2.7 |
| | PASI50 | Percentage of patients achieving 50% reduction in baseline PASI[1] at week 12. | 60.0 | 0.0 | % | 54.9 | 16.7 |
| | PGA | Percentage of patients achieving Physician's Global Assessment[2] clear/almost clear at week12. | 40.0 | 0.0 | % | 29.5 | 5.1 |
| | OLS | Percentage of patients with Overall Lesion Severity rating of minimal or clear at FT (day 84). | 40.0 | 0.0 | % | 32.1 | 2.9 |
| | DLQI | Dermatology Life Quality Index[3]. Mean percentage of patients showing an improvement. | 10.0 | 0.0 | Change score | 5.8 | 2.1 |
| Unfavourable Effects | AEs | Percentage of patients exhibiting injection site reactions, mild to moderate dose-related acute flu like symptoms. | 50.0 | 20.0 | %/100ptyrs | 41.0 | 24.0 |
| | Severe infections | Proportion of patients experiencing infections serious enough to require hospitalisation. | 3.00 | 0.00 | %/100ptyrs | 2.83 | 1.4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Severe Thrombocytopenia | Number of cases exhibiting severe (grade 3 and above) thrombocytopenia[4]. | 10 | 0 | number | 9 | 0 |
| | Psoriasis Severe Forms | Percentage of patients developing severe forms of psoriasis (erythrodermic, pustular). | 4.0 | 0.0 | % | 3.2 | 1.4 |
| | Hypersensitivity Reactions | Percentage of patients exhibiting hypersensitivity reactions, arthralgia, psoriatic arthritis, flares, back pain asthenia, ALT and Ph. Alk increase. | 10.0 | 0.0 | % | 5.0 | 0 |
| | Interstitial Lung Disease | Number of cases of interstitial lung disease. | 20 | 0 | number | 18 | 0 |
| | Inflammatory Polyradiculopathy | Number of cases of inflammatory polyradiculopathy. | 5 | 0 | Data | 4 | 0 |
| | SAEs | Number of cases of haemolytic anaemia. | 25 | 0 | number | 24 | 0 |
| | PML | Number of cases of progressive multifocal leukoencephalopathy. | 5 | 0 | number | 3 | 0 |
| | Aseptic Meningitis | Number of cases of aseptic meningitis. | 30 | 0 | number | 29 | 0 |

[1]PASI is a measure of the average redness, thickness and scaliness of the lesions (each graded on a 0-4 scale), weighted by the body region and the area affected. PASI range is from 0 to 72.

[2]PGA is a seven point scale with 7 being clear, 6 almost clear, 5 mild, 4 mild to moderate, 3 moderate, 2 moderately severe and 1 severe psoriasis.

[3]DLQI is a 10-item quality of life index scored by the patient on a four point scale.

[4]As shown in laboratory test results that indicate a decrease in number of platelets in a blood specimen.

The Hiview[2] computer program converted the scores of the drug and placebo on those measurement scales into 0-100 preference value scales. Either direct linear transformations (higher measures are more preferred) or inverse linear (lower measures are more preferred, as for mean change in PGA score). An exception was PML, for which a non-linear value function was deemed more appropriate over the whole range from 0 to 5 cases per patient year. Participants assessed the value function shown in Figure 2; this effectively captures the non-linear clinical relevance of the number of PML cases.

Weights later assigned to the criteria ensured the equality of units of the preference values on all scales. It is this conversion from different input measures into preference values, whose criterion scales are later weighted, that enable quantitative comparisons of benefits and risks.

It is apparent that some double-counting exists in the favourable effects. The proportion of patients achieving PASI75 is included in the proportion of patients PASI50. The subsequent weighting process took this into account by ensuring that the sum of weights on these two scales considered together was in the desired proportion to the other scales.

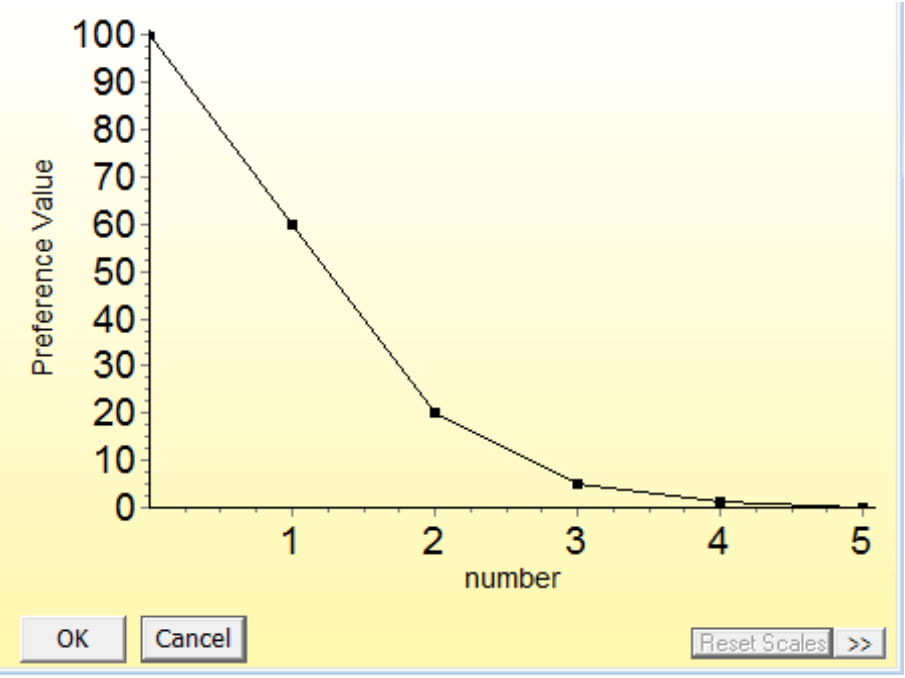



**Figure 2: The group's assessed value function for number of PML cases.**

[2] Hiview was originally developed at the London School of Economics & Political Science, and is now developed and available from Catalyze Limited, www.catalyze.co.uk.

#### *5.3.3.3 Scoring the Options*

Measures expressing the performance of the options on each criterion were determined by the group on the basis of the pooled data, and entered into the computer. An example, PML, is shown in Figure 3. Input data on the left are displayed on the thermometer scale, whose range from 0 to 5 cases encompasses the entire range of uncertainty about this effect. The right panel shows the computer's inverse linear conversion of those scores onto a 0-100 preference scale.

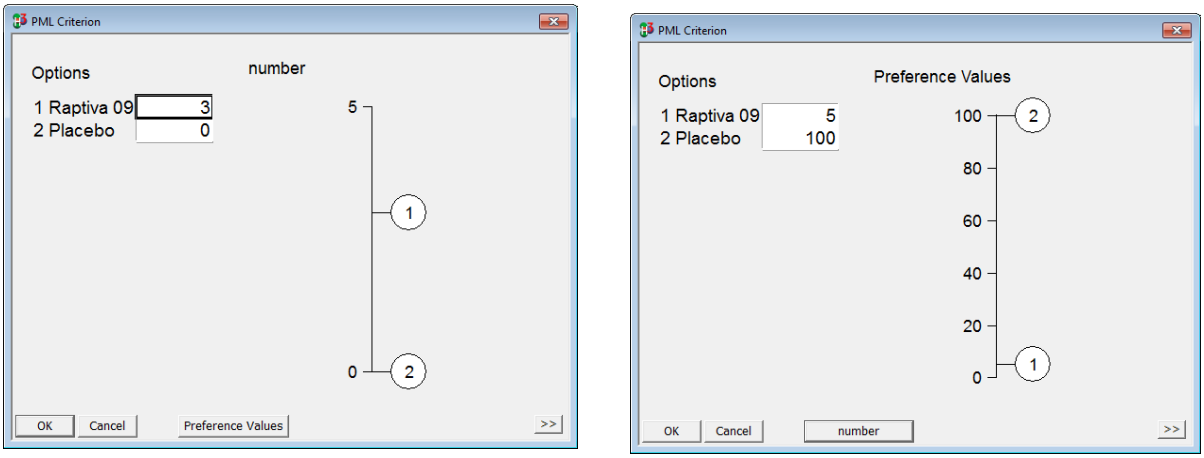



**Figure 3: Input data for the two options on the PML criterion, left panel, and their conversion into preference values, right panel, showing that lower proportions of the AE are more preferred, and that the non-linear value function, shown in Figure 2, substantially increases the difference between the drug and placebo.**

At this stage in the analysis, all input data had been converted into 0-100 preference-value scales. As there are 10 such scales, the next task was to ensure that the units of preference value were equivalent across all the scales. That is the purpose of weighting.

#### *5.3.3.4 Weighting*

Some criteria are more clinically relevant expressions of preference value than others. Although that is an intuitively appealing statement, more precision is needed to enable the assessment of weights for the criteria. To ensure that assessed weights are meaningful, the concept of 'swing weighting' was applied. As an analogy, both Fahrenheit and Celsius scales contain 0 to 100 portions, but the swing in temperature from 0 to 100 on the Fahrenheit scale is, of course, a smaller swing in temperature than 0 to 100 on a Celsius scale; it takes 5 Celsius units to equal 9 Fahrenheit units. The purpose of weighting in decision theory is to ensure that the units of preference value on the

different scales are equivalent, thus enabling weighted scores to be compared and combined across the criteria. Weights are, in essence, scale factors.

It follows, then, that to judge preference value, two steps in thinking must be separated. First, it is necessary to think about the difference in the measured effect represented by a preference value of 0, compared to the level of effect represented by a preference value score of 100. That is a straightforward assessment of a difference in effect, from the least preferred effect to the most preferred effect on that criterion. The next step is to think about how much that difference in effect matters; this is essentially a judgement of the clinical relevance of the difference in effect size. "How big is the difference and how much do you care about that difference?" This is the question that was posed in comparing the 0-to-100 swing in effect on one scale with the 0-to-100 swing on another scale.

During the decision conference participants first assessed weights within each right-most grouping of favourable effects, the four Physicians' ratings criteria first. Figure 4 shows the weights for that grouping. The group agreed that the swing from 0% to 60% on the PASI75 scale was better than any of the other improvements, so the PASI75 was assigned a weight of 100. Compared to that, the group judged the swing on the PGA scale to be nearly as good, and agreed a weight of 80.

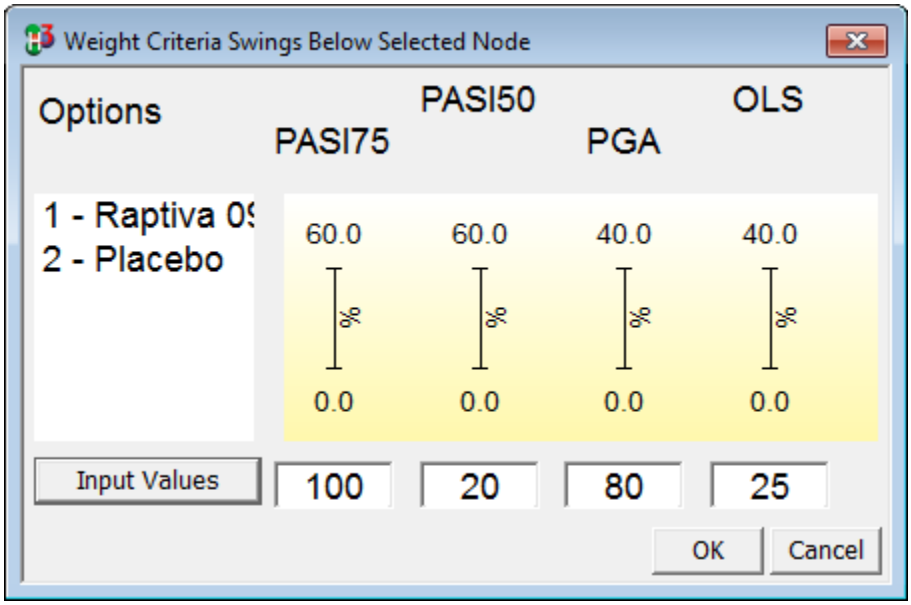



**Figure 4: The swing-weights assigned to the four Physicians' ratings scales.**

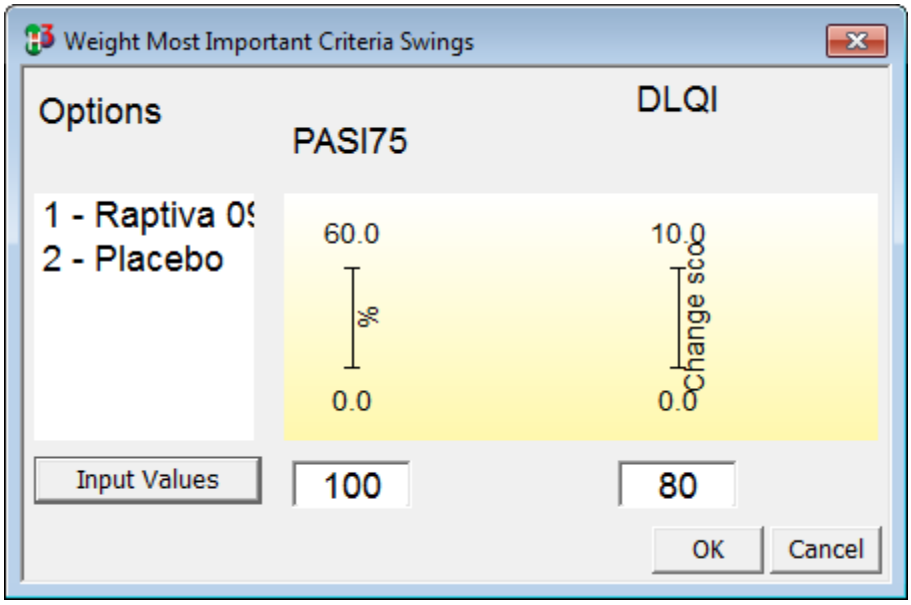



**Figure 5: Swing weights assigned to the 100-weighted criteria for PASI75 and PGA**

In the next step, the group compared the PASI75 scale with the DLQI scale, assigning the latter a weight of 80 compared to the PASI75, as shown in Figure 5.

The group then turned to weighting the Unfavourable Effect criteria, starting with the SAE criteria; the largest swing weight was judged to be for Serious Infections, so that criterion was given a weight of 100. Next, that criterion was compared to AE, which was assigned a weight of 20. Then, moving to the criteria under Observational data, the group quickly agreed that the 0-to-5 swing for PML was the most important, so it was given a weight of 100, and the other swings were judged relative to that 100. Comparing the 100-rated swing under Clinical Trials, Serious Infections, with the 100-rated swing under Observational data, PML, resulted in an assessed weight of 20 for Serious Infections compared to the 100 for PML.

The final, and most difficult comparison, is shown in Figure 6: PASI75 versus PML. After considerable debate, the group agreed that the PML swing, from 5 cases down to none, was half the clinical relevance of PASI75, from 0% to 60% of patients achieving PASI75. But sensitivity analysis on that weight was promised, for not everybody agreed that 2 to 1 was the final answer.

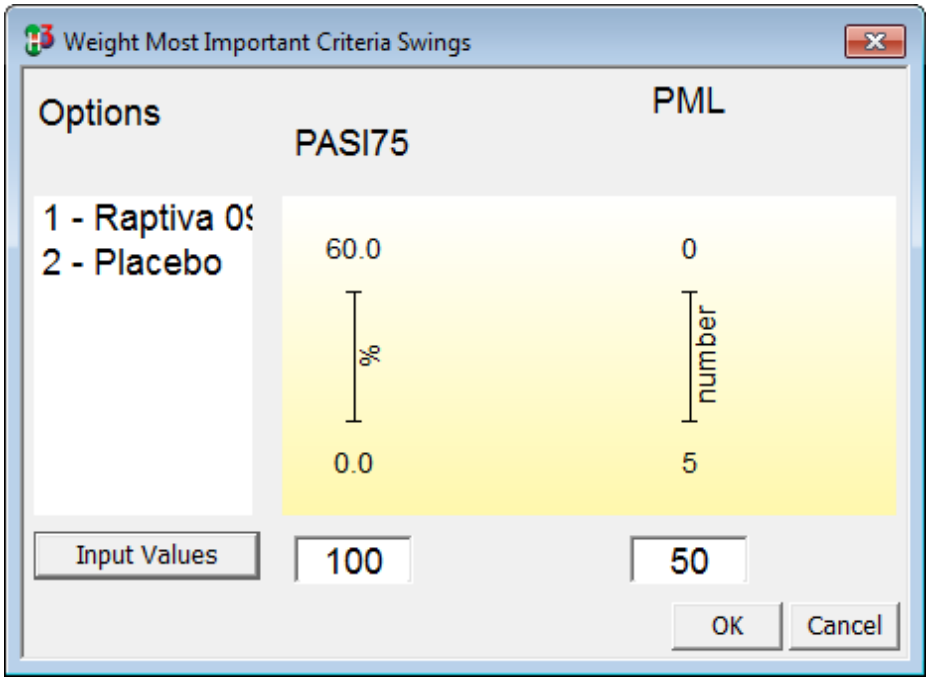



**Figure 6: Swing weights comparing PASI75 to PML.**

It is this process of comparing swings from least to most preferred positions on the criteria associated with a node, assigning one criterion swing a weight of 100, then comparing the 100-weighted criteria across the nodes, which ensures the comparability of the units of preference values across all the criteria.

It is easy to become lost in attempting to understand the weighting process by reading about it, so Figure 7 shows all the originally-assessed weights, each divided by 100, on the value tree. Hiview multiplies these weights along each path through the tree, sums the products for all 11 criteria and divides each product by the sum. This gives the cumulative weights shown in Figure 10, re-normalised to 100, with the criteria sorted in order of the cumulative weights.

It is important to keep in mind that a cumulative weight represents the total added preference value in moving from the least to most preferred positions on a scale. These weights represent the relative importance of the 0-100 preference value ranges on the scales, not the relative importance of favourable and unfavourable effects, and particularly not the relative importance of those effects for the drug and placebo. By summing cumulative weights, it is possible to see the weights at each node. For example, the sum of all the favourable effects weights is 78 with 22 for the unfavourable

effects. In other words, the total range of 0-100 differences in preference values on the favourable effects three-and-a-half times the range of that on the unfavourable effects.

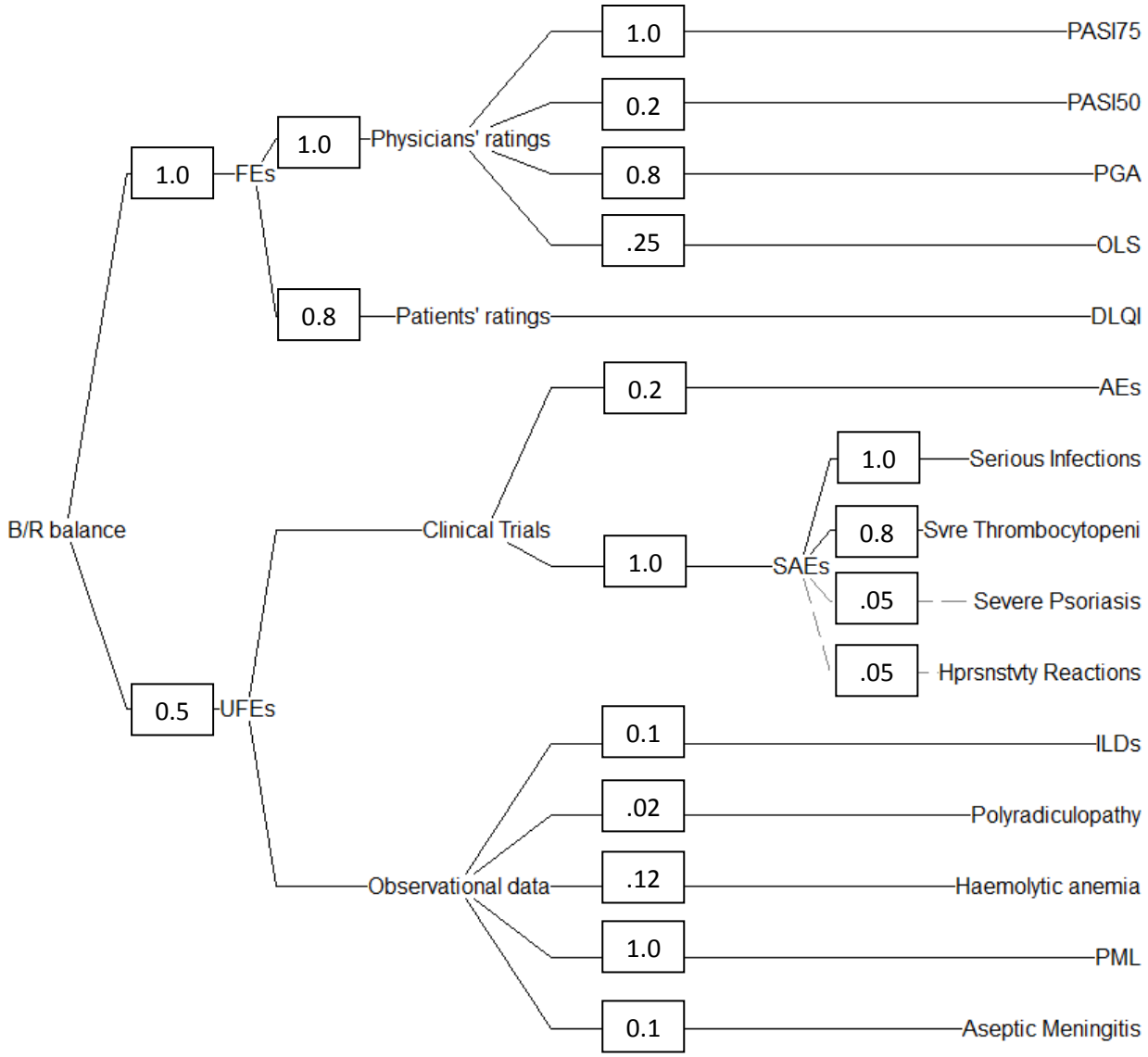



**Figure 7: The originally-assessed swing-weights, divided by 100, assigned at all the nodes.**

### 5.3.4 Results

#### *5.3.4.1 Overall*

With scoring and weighting completed, it was possible to calculate sums of weighted preference values and show preliminary results at any node. Figure 9 shows the relative scores at the FE/UFE Balance node of Figure 1 as stacked bar graphs. Each section of each bar graph shows the contribution of favourable effects and unfavourable effects to the overall score, which is shown at

the bottom of the bar. Note that longer green bars represent *more* benefit, while longer red bars represent *more safety*. Efalizumab shows a 20-point advantage over the placebo.

The stacked bar graphs can also be shown for their separate contributions from the criteria, as seen in Figure 10. This instructive display shows the three main advantages of efalizumab: PASI75

PGA and DLQI. Collectively, they far outweigh the advantages of the placebo: its modest side effects and absence of PML. However, as the group learned, this result depends on the relative weights between the favourable and unfavourable effects, explored below.

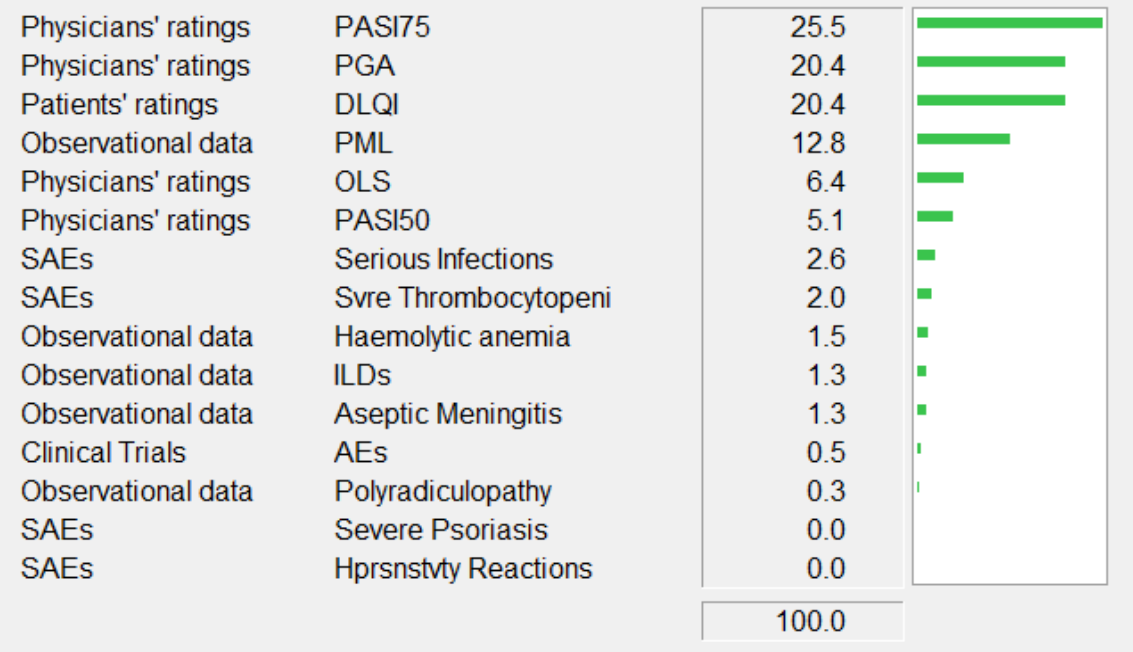



**Figure 8: Cumulative weights of all the criteria, with the criteria ordered by the size of their cumulative weights, which represent the swings in preference from the least to the most preferred positions on the scales.**

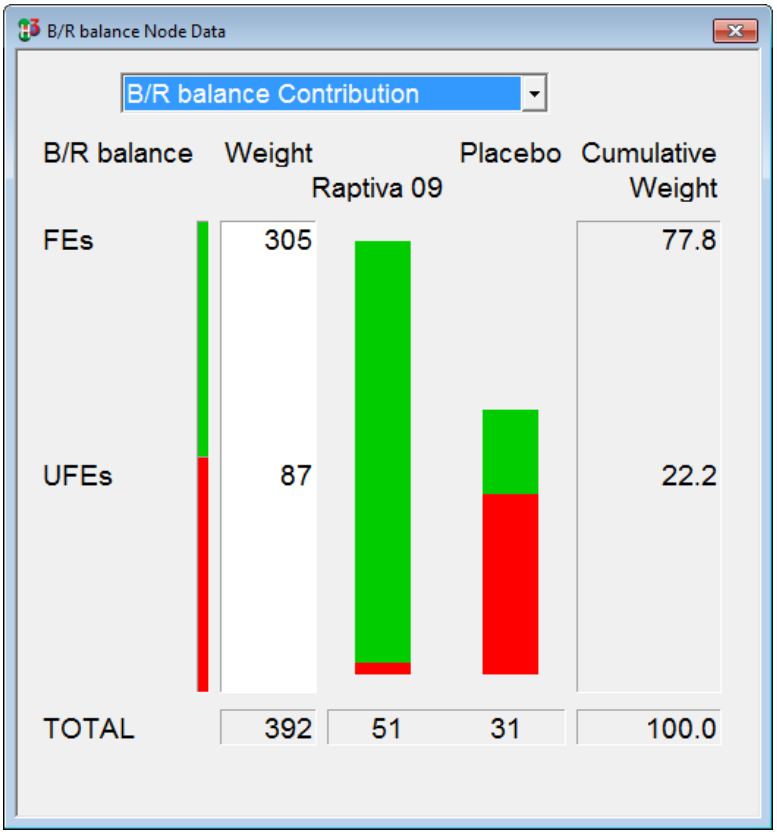



**Figure 9: Overall Benefit-Risk balance for** efalizumab**. Longer green bars represent more benefit, while longer red bars show more safety. The Cumulative Weight column shows the normalised weight on the FE and UFE nodes, favourable effects weighted more than three times as much as for unfavourable effects.**

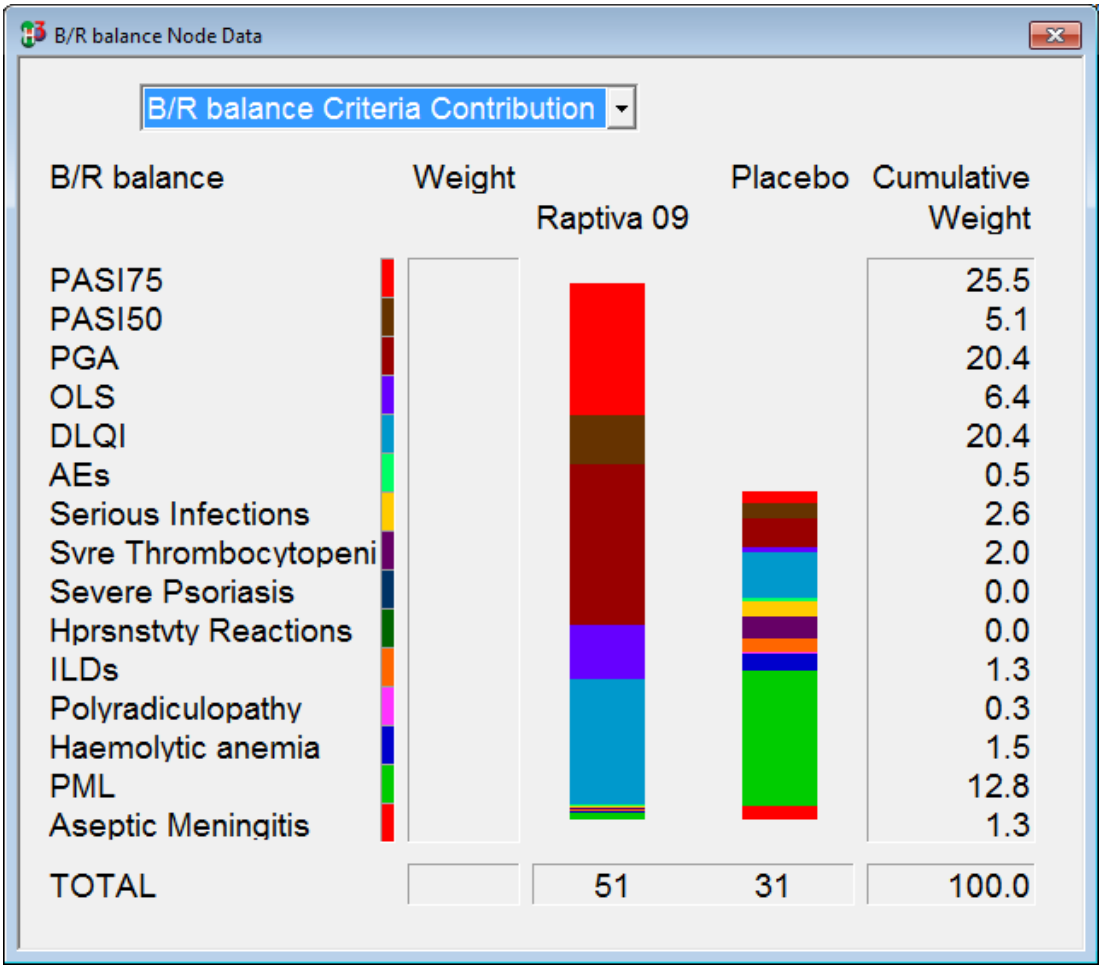



**Figure 10: The drugs ordered by their overall weighted preference scores, with the stacked bar graphs showing the contribution to the overall score of the criteria. The right column shows the cumulative weights, normalised to 100, of each of the criteria. Flare rate, for example, is 20.2.**

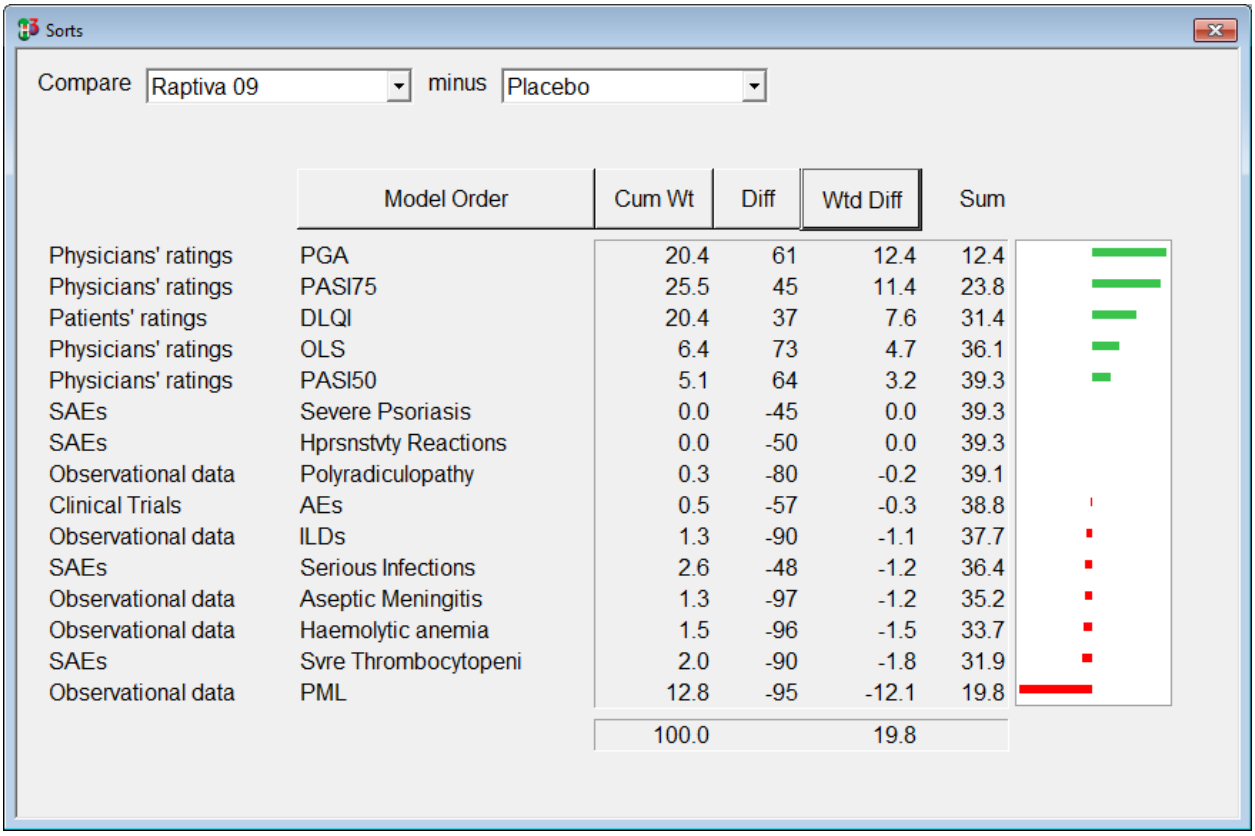
#### *5.3.4.2 Comparative Analyses*

A more clear display of the differences between efalizumab and the placebo can be seen in Figure 11. The Diff column in each display shows the difference in the preference scores, while the Wtd Diff column multiplies that difference by the cumulative weight on the criterion. It is this weighted difference that reveals the true advantages and disadvantages of the comparisons, criterion-by-criterion. They are the 'part scores', whose sum, 19.8, represents the overall weighted difference of preference values for the two options.

The two main advantages of efalizumab are PGA and PASI75. Note that the PASI50, the primary endpoint, is in fifth position. It shows a large preference-value difference of 60 compared to the placebo, but the weight on that criterion is a quarter as large as the weight on PASI75. For the latter, the difference score of 45 is smaller, but that criterion is more heavily weighted, so the

weighted difference score on PASI75 of 11.4 is nearly four times as large as the weighted preference score on PASI50.

Although the efalizumab-Placebo difference for patients' ratings, DLQI, is the smallest of the favourable effects at 37 points, it is on a heavily-weighted criterion, with the result that the weighted difference score is more than twice that of the primary endpoint.




| | Model Order | Cum Wt | Diff | Wtd Diff | Sum |
|---|---|---|---|---|---|
| Physicians' ratings | PGA | 20.4 | 61 | 12.4 | 12.4 |
| Physicians' ratings | PASI75 | 25.5 | 45 | 11.4 | 23.8 |
| Patients' ratings | DLQI | 20.4 | 37 | 7.6 | 31.4 |
| Physicians' ratings | OLS | 6.4 | 73 | 4.7 | 36.1 |
| Physicians' ratings | PASI50 | 5.1 | 64 | 3.2 | 39.3 |
| SAEs | Severe Psoriasis | 0.0 | -45 | 0.0 | 39.3 |
| SAEs | Hprsnstvty Reactions | 0.0 | -50 | 0.0 | 39.3 |
| Observational data | Polyradiculopathy | 0.3 | -80 | -0.2 | 39.1 |
| Clinical Trials | AEs | 0.5 | -57 | -0.3 | 38.8 |
| Observational data | ILDs | 1.3 | -90 | -1.1 | 37.7 |
| SAEs | Serious Infections | 2.6 | -48 | -1.2 | 36.4 |
| Observational data | Aseptic Meningitis | 1.3 | -97 | -1.2 | 35.2 |
| Observational data | Haemolytic anemia | 1.5 | -96 | -1.5 | 33.7 |
| SAEs | Svre Thrombocytopeni | 2.0 | -90 | -1.8 | 31.9 |
| Observational data | PML | 12.8 | -95 | -12.1 | 19.8 |
| | | 100.0 | | 19.8 | |

**Figure 12: Efalizumab compared to the placebo. The sum of the five favourable effects, 39.3, outweighs the sum of the unfavourable effects, 19.5, to give an overall weighted preference value of 19.8 in favour of efalizumab over the placebo.**

#### *5.3.4.3 Sensitivity Analyses*

These analyses explore the sensitivity of the overall results to changes in weights on the criteria, which were the source of much of the debate about the balance of benefits and risks. The first analysis examined the weight on the unfavourable effects to see if increasing that weight, and thereby decreasing the weight on the favourable effects (so that the total cumulative weights continue to sum to 100) would tip the benefit-risk balance in favour of the placebo. The normalised weight on the Unfavourable Effects node was 22.2, as shown in the right column of Figure 9. The

computer varied that weight over its entire feasible range, 0 to 100, with the result shown in Figure 13.

The vertical red line intersects the horizontal axis at 22.2, and its intersections with the red and green lines give the overall scores for the efalizumab doses and the placebo, 31 and 51. Increasing the weight on the UFEs node increases the overall preference scores for the placebo and decreases the score for the drug. Increasing the cumulative weight to about 37 changes the most preferred option from efalizumab to the placebo, at the intersection of the two lines and indicated by the transition in background colour.

Brief experimentation with the relative swing weights on PASI75 compared to PML reveals that the two overall weighted scores on the two options are 43 for efalizumab and 42 for the placebo when the weights shown in Figure 5 are 100-100, i.e., 60% of patients experiencing a 75% reduction in baseline PASI is as clinically desirable as 5 cases of PML is undesirable. This can be seen graphically in Figure 14.


**Figure 13: Increasing the weight on the UFE node from its current value of 24.1 shows that the weight would have to more than double for the placebo to be preferred.**

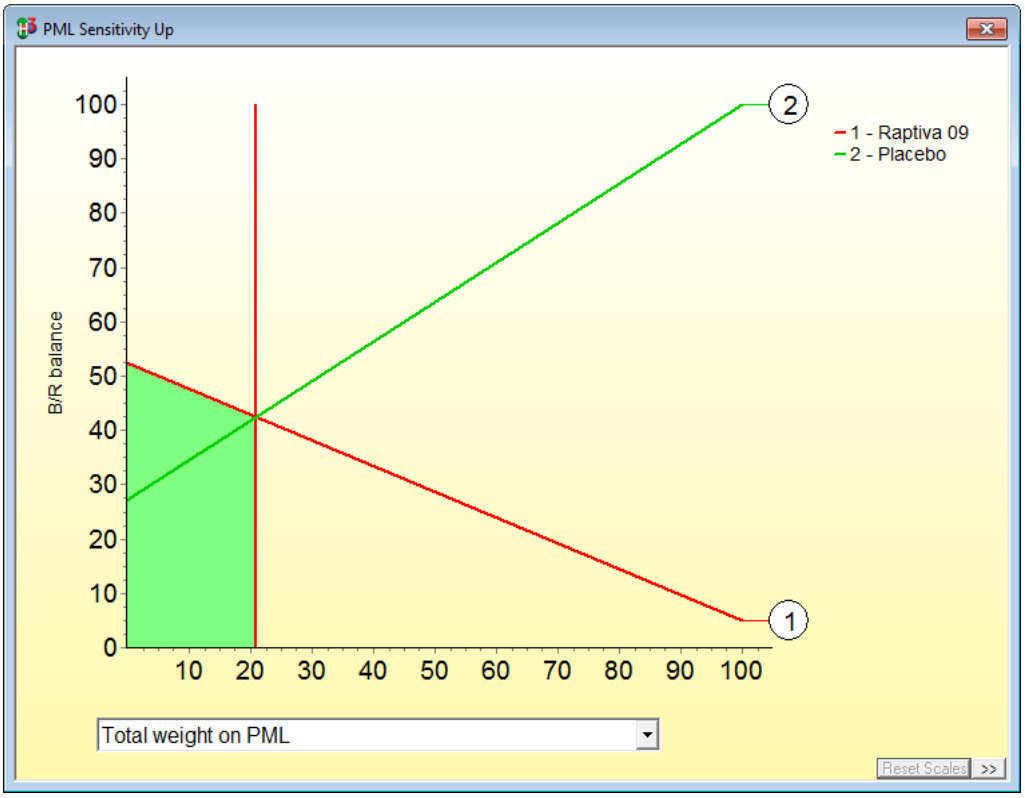



**Figure 14: Increasing the weight on PML to equal that on PASI75 shows that equal clinical concern for these two effects results in equal overall weighted scores for efalizumab and the placebo.**

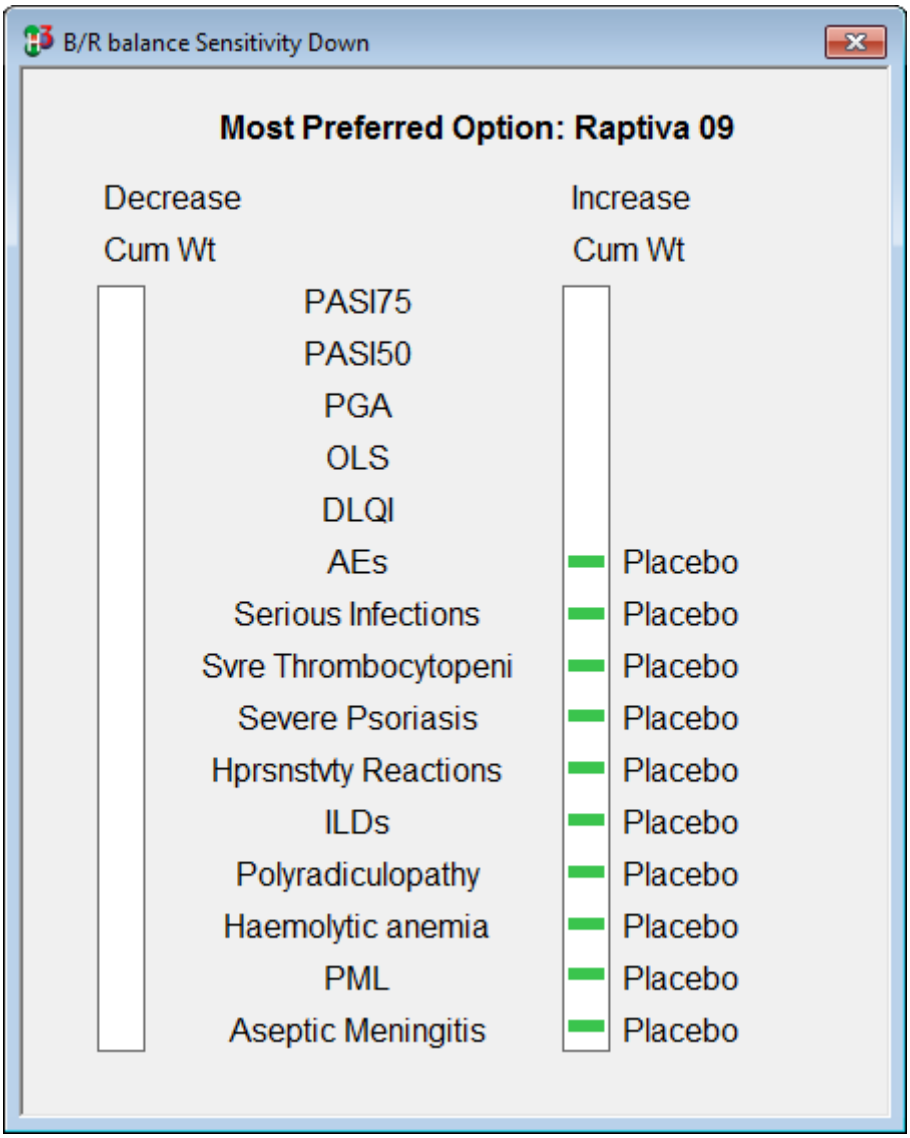



**Figure 15: Separate sensitivity analyses on each of the criteria shows how the most preferred option, the efalizumab, would change as the cumulative weight on a criterion is decreased or increased. Green bars show cumulative weight changes greater than 15 points are needed to shift the overall preferences. Had a yellow bar appeared, it would signal a change of 5 to 15 points would change the result, while a red bar would indicate that a small change in a weight, less than 5 points, would change the most preferred option. Here, the absence of any bars for the five favourable effects, and no yellow and red bars indicates a robust model.**

After returning the relative weights on PASI75 and PML to their base-case values of 100 and 50, the group explored whether or not there were any more crucial judgements that could shift the results. A simultaneous sensitivity analysis on all the criteria indicates which criterion weights make a difference. Figure 15 shows the summary display, with efalizumab at the top as the most preferred

option. The middle column lists the criteria, while the right column shows the results of increasing the cumulative weight on each criterion independently, and the left column the result of decreasing the cumulative weight.

As noted in the previous analysis, PML just barely missed a yellow bar, but that really is the only sensitive criterion. The weight on any single unfavourable effect has to be increased substantially to change the overall result, while changing the weight on any single favourable effect, increasing or decreasing it, will not by itself push the placebo into first place.

### 5.3.5 Discussion and Conclusions

The overall result of the modelling showed that the benefit-risk of efalizumab is substantially better than that of the placebo, even taking into account the three PML cases. This conclusion is robust to substantial differences of opinion about the individual weights on the criteria. Indeed, orders of magnitude increases would be required for the unfavourable effects, except for PML, to tip the balance. Only when more weight is given to 5 cases of PML compared to 60% of patients achieving a 75% reduction in baseline PASI would the model favour the placebo over efalizumab.

So, why did the CHMP recommend in February 2009 that marketing authorisation for efalizumab should be suspended? The official public statement explains that "its benefits in the treatment of psoriasis were modest, while there was a risk of serious side effects, including the occurrence of progressive multifocal leukoencephalopathy (PML)". The suspension could be lifted if a sub-population could be identified for whom the benefits would outweigh the risks. The Marketing Authorisation Holder declined to conduct the necessary clinical trials, so the European Commission withdrew marketing authorisation for efalizumab in June 2009.

Is there a conflict between the decision of the CHMP and the model results reported here? The answer is "not necessarily". Models don't make decisions; people do. Models simply reflect back, in changed form, the information given to them. For the efalizumab model, the information provided includes the criteria shown in the Effects Tree, the measured data from the clinical trials and the incidences of unfavourable effects from the post-authorisation period, the judgement of the value function for PML and the assessments of swing-weights for the criteria. The pooled information on which the model results are based does not necessarily reflect all the available information, for the data are not always reported fully in the publicly-available reports. It is difficult to reconstruct today what was in the minds of assessors in 2004, 2008 and 2009, what data they used and how they pooled the available information. Modelling is best done at the time when a recommendation is required and the issues are 'hot'. Thus, a shortcoming of the model reported here is that it may not adequately reflect the situation experienced by assessors in early 2009.

By 2009, information in addition to the clinical studies had become available, but it is difficult to determine from the public assessment reports what new information led to the view that efalizumab's benefits were "modest". Indeed, the April 2008 Assessment Report for efalizumab (EMEA/112794/2009) notes that for Study 25300 "the response rate [PASI75] in patients (n=232) who were refractory to all three major systemic treatments (i.e. cyclosporine, methotrexate, and PUVA) was 61% versus 69% in patients not refractory for any of these (p=0.03)". From the perspective of the patient who was unresponsive to the other treatments, this is not a modest effect.

But the reporting raises the issue of what is mean by a 'modest effect'. That phrase first appears in the EPAR, on page 36, as a summary of the finding that 27% of patients achieved PASI 75 (the primary endpoint). Data reported in the Effects Table in this report show similar percentages of patients achieving some sort of improvement, judged by physicians or patients. All the percentages shown there are around 30% (except for the PASI 50, which is generally disregarded in the Assessment Reports as being of little clinical significance). It would appear that 'modest' is more a public health interpretation, in that less than one third of psoriasis sufferers would be helped, than it is an indication that the efficacy itself will be modest. In other words, a psoriasis patient reading the EPAR might conclude that he or she would only experience modest relief, when in fact the data show that for responders the efficacy could be considerable.

In short, the public health perspective of regulators can lead to potential communication problems for failing to distinguish between the magnitude of an effect from an individual's uncertainty that they will benefit from the effect. It might have been clearer to report that "27% of patients can expect to experience a 75% reduction in their condition".

Returning now to the question of whether the efalizumab model conflicts with the CHMP's final recommendation to withdraw the product, it is important to recognise that the function of a decision model is to serve as a 'tool for thinking', a decision aid that provides as many answers as there are judgements and assumptions provided as inputs. Many answers arise from disagreement about inputs. Experience of modelling five drugs during the EMA's Benefit-Risk Project, and more generally of working with teams of stakeholders and key players, shows that experts and assessors frequently disagree. Bringing them together in groups allows them to share their differing perspectives and experience so that informed assumptions, judgments and assessments can be tested for their effects on the overall benefit-risk balance, as described in Appendix A in the "Supplement 2 to Wave 1 case study report Efalizumab" (http://www.imi-protect.eu/benefitsRep.shtml). Thus, a model gives as many different results as there are different inputs, but the process will enable the assessors to achieve a shared understanding of the important factors that affect the benefit-risk balance, to develop a sense of whether the benefit-risk balance is favourable or unfavourable, and, finally, agreement about what recommendations to make. Consensus about inputs is not required to achieve this level of agreement about the way forward.

# 6 Discussion

## 6.1 Methodology

### 6.1.1 Assessment of appropriate frame for benefit-risk approaches through practical experience

| | Comments | Proposed improvements and/or extensions |
|---|---|---|
| PrOACT-URL | PrOACT URL was applied as it provided a strong detailed context to our case study. It helped to evaluate which further methodologies could, or could not be implemented based on the available data, aims and objectives.<br>Implementation was completed with ease. Technical demand was not taxing. However, the description of steps was a little unclear in places. Although the brevity of instructions was beneficial for fast progression, occasionally they lacked clarity.<br>• E.g. 1d."affected population" could refer to psoriasis patients, or patients receiving efalizumab<br>The anticipated time to be spent on the framework was not so much determined by the framework itself, and instead on the amount of literature necessary to read and extract data.<br>The steps are laid out in point form, which clearly delineate which fields are to be completed. The framework places a strong emphasis upon previously collected data and sources, ensuring the inclusion of relevant regulatory literature, data on favourable effects, and data on unfavourable effects.<br>Creative, doable alternatives are investigated in Step 5 Alternatives, where options to be evaluated against criteria are identified.<br>The method can apply to any stage of a product lifecycle, i.e. early development to post-marketing, where decision-making may occur to provide an explicit statement of objective, context, benefits and risks. | To provide a user guide which explains with greater clarity which information is being requested, and the rationale for documenting it. Some strong worked examples of how PrOACT-URL should be applied would be good. |

| | | |
|---|---|---|
| BRAT | The approach was chosen so as to provide a comparator against PrOACT-URL. This is because similarly to PrOACT-URL, it also provides a strong detailed context for the case study through the selection, inclusion and presentation of favourable and unfavourable effects.<br>Technical demand is necessary via statistical knowledge. For our case study, it was necessary to perform a meta-analysis of beneficial effects, and also calculate relative risks and risk differences (with confidence intervals).<br>The instructions are laid out within an extensive user guide with worked examples, and provide a strong and comprehensive description of methodological steps.<br>While it offers a strong emphasis on data and sources, it also offers the additional benefit of providing an explicit audit trail, allowing for a high degree of transparency.<br>PhRMA BRAT Softpilot software was used and relatively intuitive to learn. However, there were small glitches with the beta version of software e.g. red and green colouring of favourable and unfavourable effects was not automatic. Also, one comparator at a time can be used, although it is relatively easy to switch between the fields. Currently the software can only handle dichotomous variables which will limit its applicability in situations where continuous variables are important e.g. oncology.<br>Similarly to PrOACT URL, the method can apply to any stage of a product lifecycle, i.e. early development to post-marketing, where decision-making may occur to provide an explicit statement of objective, context, favourable and unfavourable effects.<br>Creative, doable alternatives are essential for this methodology. The method is only possible when a comparator such as placebo, background epidemiological rates, or active comparator. This is because relative risks and risk differences have to be calculated in order for visual representation on the forest plot, or inclusion in the Key Benefit-Risk Summary Table. This may present a limitation for selecting benefit-risk criteria because data must be available for a comparator. | Include a section for suggested formulae for calculating relative risks and risk differences (with confidence intervals). |

| | | |
|---|---|---|
| BRR | BRR is a method when the benefit-risk analysis involves one benefit criterion and one risk criterion. In case there are multiple benefits or multiple risks, either multiple criteria are collapsed into one criterion, or a primary benefit and a primary risk are selected and ignore the others. | |
| MCDA | Qualitative frameworks do not provide an integrated B-R assessment unless supplemented with a quantitative method, however provide a clear and structured approach to problem to be resolved.<br>Few methods allow for the assessment of B-R with multiple criteria and multiple options, although this is closer to real-life situations.<br>No method provides clear guidance for assessing Benefit and Risk when the timeframe for outcomes is significantly different and when the outcome depends on time. In the efalizumab case study the main FE was measured at 12 weeks and the main UFE was measured at year 4.<br>Of note, the Problem was reframed from the initial mandate given to CHMP (maintain, vary, suspend or revoke the Market Authorisation) to "placebo 2004 versus efalizumab 2009"; this allowed to take into account all the observational data which had an impact on the Risk of efalizumab after 4 years on the market; however it didn't take into account that, in 2009, the context was different with 3 new competitors with a more favourable efficacy, hence decreasing the relative Benefit of efalizumab as compared to these competitors. | |

### 6.1.2 Assessment of using meaningful reliable information for benefit-risk approaches through practical experience

| | Comments | Proposed improvements and/or extensions |
|---|---|---|
| PrOACT URL | The rationale for including or excluding benefit and risk criteria was to initially include every favourable and unfavourable measure from clinical trials and post-marketing surveillance for which data was publically available (e.g. EPARS, PSUR10, Scientific Discussion). This was to ensure that all of the | |

| | | |
|---|---|---|
| | evidence available to the regulator was accounted for in order to provide a strong context for decision-making.<br>Clinical judgements about the effects were available, and considered within Steps 7 to 11.<br>MCDA is recommended. | |
| BRAT | Similarly to PrOACT URL, the rationale for including or excluding benefit and risk criteria was to initially include every benefit and risk measure taken from clinical trials and post-marketing surveillance. Next, all effects with non numerical or missing data (e.g. no background epidemiology or placebo/comparator data) were excluded. From the remaining measures, it is possible to place filters on specific branches and change the value tree according to perception of stakeholder preferences. Major limitation for post-marketing data based on spontaneous reporting (no incidence) without background incidence rates.<br>PhRMA BRAT can account for criteria other than efficacy and safety,<br>The framework can be used to present qualitative information. However, this option is not available when using the software, and qualitative information does not appear in the forest plot or Key Benefit-Risk Summary Table.<br>Clinical judgements about the effects are considered in Step 5 Assess Outcome Importance, where values are assigned according to the perspective of the decision-maker. Simple methods of weighting, e.g. categories of importance, ranking, ad hoc weights, and direct assessment/point allocation are suggested. There is also the suggestion of more complex weighting via MCDA and conjoint analysis. | There were a significant number of effects without a comparator. These were excluded primarily because the necessary due diligence to identify the comparator data in the literature, other databases was not possible because of time limitations they could not be represented by the software, or resulting visualisations. It would be good to have some specified contingency plans in the event that no comparator is possible.<br><br>There could be an expansion on the software to represent additional formats of data other than numerical. |
| MCDA | "Publicly available" information does not necessarily include the source data which would allow a precise measure for an outcome:<br>Meta-analyses of CTs is not always possible (different enrolled populations)<br>Format of Public documents (EPAR and scientific discussion) may not ease extraction of data.<br>Extreme heterogeneity of measures (absolute numbers, proportions with various denominators) | |

| | | |
|---|---|---|
| | is manageable in MCDA, which is useful in a post-marketing evaluation where measures are very heterogenic in nature and in units.<br>Comparators data used for the assessment are not necessary relevant for decision making in a practical medical context (e.g. placebo instead of active treatments). This depends on the Problem and decision context chosen in the framework.<br>Several measures for a same medical outcome may be redundant and lead to double counting (e.g. PASI 75 and PASI 50)<br>Scoring and weighting are very sensitive to assessor's background, experience, and conditions of the B-R assessment (emergency or delayed)<br>The weight given to some outcome would deserve thorough discussion on their medical relevance (e.g. reversibility of serious risks, long term continuation of short term benefit, etc.)<br>Choice of outcomes for FE and UFE may be difficult (exhaustive list of outcomes or selection), based on which medical relevance?)<br>When measures are missing, outcome (either FE or UFE) may not be taken into account and bias the final assessment.<br>The measures made on a CT population may not reflect a "real world" population (with off label use, misuse) in a post-marketing setting B-R assessment. | |

### 6.1.3 Assessment of the availability of clear values and trade-offs for benefit-risk approaches through practical experience

| | Comments | Proposed improvements and/or extensions |
|---|---|---|
| PrOACT-URL | Unfavourable and favourable effects are defined clearly in the approach, by use of an effects table.<br>There was no common scale, as the literature presented data in many different units, e.g. %, %/100py, etc.<br>It is important to note that the framework cannot be used alone although it provides a good contextual basis to use in conjunction with other benefit-risk methodology. | |

| | | |
|---|---|---|
| | The results are easily interpretable; however, the responses must fit within limited table space. A fuller discussion may be possible e.g. 1d. "Patients' and physicians' concerns", and desirable under certain circumstances as they may significantly frame objectives. Some of the richness of discussion may be neglected in the brief description within the table. | |
| BRAT | Unfavourable and favourable effects were clearly defined, in addition to a detailed audit trail stating how effects were selected and included or excluded during specific steps. Again, there was no common scale as the regulatory documentation presented data in different units.<br>Similar to PrOACT URL, the framework requires use in conjunction with another benefit-risk methodology.<br>The final results are interpretable, but may be challenging for specific groups of stakeholders. E.g. interpreting confidence intervals, and odds ratio on a log scale. However, risk differences are presented in natural frequencies (with a denominator or 1000) which are regarded as an effective way of presenting risks to patients and the public.<br>Trade-offs are not made explicit in quantitative form. Some users of BRAT have assessed criterion weights as the importance of the criterion (though not as swing weights, as required in MCDA). But even then aggregation is not done explicitly in BRAT, and would be difficult when metrics differ. | For our case study, the denominator (even if 10,000) was not large enough to specify the risk in the Key Benefit-Risk Summary table and rounded it to 0. The framework should consider how to include rare but serious adverse events. |
| BRR | The BRR is also often used together with probability simulations. In simulation approach, both $\Delta B$ and $\Delta R$ are taken as random variables following certain distributions. The probability that $(\Delta R, \Delta B)$ falls over the threshold is calculated and is taken as evidence for decision making. | |
| MCDA | Clear values may be missing although potentially medically relevant for a final decision (e.g. efficiency of Risk Minimisation actions objectively measured by impact on some outcomes, or comparators post-marketing safety data)<br>Balancing a very rare serious effect with a | |

| | | |
|---|---|---|
| | relatively "modest" benefit may be challenging, even more so if the background of the serious effect is not nil.<br>The preference elicitation was done rather roughly and quickly despite the large representativity of the Task Force; the preference elicitation was not systematically made with the chosen perspective (regulator) but as a mix of a prescriber's, patient's and regulator's perspective. | Given the importance given to preferences besides data in the Decision models, structured and validated questionnaires should be developed and used for these methods. |

### 6.1.4 Assessment of the logically correct reasoning for benefit-risk approaches through practical experience

| | Comments | Proposed improvements and/or extensions |
|---|---|---|
| PrOACT-URL | Whether the approach can handle different forms of data, e.g. qualitative, continuous etc. depends on which method is selected for Step 9. Uncertainty. | |
| BRAT | The approach can handle many different forms of data including, qualitative, quantitative, objective and subjective. However, if the software is to be used, only numeric data can be represented.<br>How uncertainty is accounted for is dependent upon confidence intervals, and the weighting method selected in Step 5 Assess outcome importance.<br>Effects are not combined in the forest plot and Key Benefit-risk Summary Table. | Consider how continuous data could be incorporated, e.g. mean increase in DLQI point scoring since baseline. |
| BRR | BRR decision depends on the threshold choices. Threshold line implies that the tolerance of risk increase is proportional to the benefit increase. In reality the tolerance of risk increase may not be linear to the benefit increase. In this situation a threshold curve can be used instead of a straight line. The threshold elicitation can be conducted by either a decision conference or a properly designed questionnaire. | |
| MCDA | The model is not supposed to provide a decision per se, but as a help to decision-making through transparency of criteria and preference | |

| | | |
|---|---|---|
| | elicitation; however this transparency is not total if some criteria cannot be included in the model because unmeasurable | |

### 6.1.5 Commitment to action

**Table 5 Assessment of the commitment to action for benefit-risk approaches through practical experience**

| | Comments | Proposed improvements and/or extensions |
|---|---|---|
| PrOACT-URL | PrOACT URL provides insight by providing a strong context to decision-making with a transparent framework. The applicability of the final results to the decision to be taken is limited by the secondary method used within Steps 8 and 9, Uncertainty. The brevity of the reporting space within the table limits the inclusion of an audit trail. | |
| BRAT | The approach provides insight by visually representing each unfavourable and favourable effect in isolation, while providing confidence intervals. The results are easily communicable, and highly transparent due to the documentation of an in-depth audit trail for every step. The value tree, Key Benefit-Risk Summary Table, and forest plot are easily exportable from the software into Microsoft Word and PowerPoint. Unfavourable and favourable effects are not combined into a common metric. In order to derive a final decision from the final results, the unfavourable and favourable effects must be combined. | |
| MCDA | The method provides a combined Benefit and Risk evaluation, which can subsequently be used for Sensitivity analysis, allowing for several possible decisions in situations where the balance would be sensitive to various scenarios. The MCDA model is also applicable where there are few or no objective measures but only preferences. | |

## 6.2 The assessment of benefit-risk balance

### 6.2.1 Benefit-risk of efalizumab versus placebo

PrOACT-URL didn't provide an explicit B-R assessment. In contrary it helped structuring a very complex situation and problem solving issue, with some important missing data, very different timeframes for outcomes, an accumulation of new information over time, a changing comparative environment over time, etc.

The BRAT framework provided a very clear display of both benefits and risks, but didn't go to the point of an integrated B-R assessment. This is completed by a simple BRR applied to the most prominent Favourable Effect and the most Unfavourable Effect. In this framework using this model, efalizumab B-R was negative in the frame of the chosen options, but a sensitivity analysis would provide a more ambiguous outcome depending on more aggressive hypothesis (larger exposure at > 3years and different threshold).

The MCDA analysis, through a well conducted Decision Conference, came to a conclusion which was actually contradictory with the one it was supposed to replicate with the same data set. Sensitivity analysis didn't provide a clear explanation for this contradiction, showing that decision makers may also include in their decision criteria which are not explicit or are not measurable. However the framework chosen didn't take into account the changed competitors' context, leaving the B-R analysis relatively isolated from this latter. The existence of some more efficient competitors at the time of the CHMP re-evaluation of efalizumab would have substantially decreased the relative Benefit of efalizumab as compared to these competitors.

## 6.3 Visual representation of benefit-risk assessment results

PrOACT-URL does not come with any specific Visual representation of its content.

BRAT was used with its proposed Forest Plot with confidence intervals, which is an easily understandable presentation for specialists.

MCDA is applied in the Hiview3 software with several graphical representations (Effects Tree, various colored bar graphs etc.) which provide easily understandable visualization of results. This is easily provided by the software itself instead of requiring an initial excel sheet.

# 7 Conclusion

The experience of the efalizumab Task Force lead to several observations which may be of interest for the choice and the use of models in a wave 2 of case studies, or even when applying some of the lessons learned in this one.

Overall, although based on the same data set, the application of two frameworks used with two different quantitative methods lead to divergent results within the applied models, as well as with the historical similar decision made in 2009 by the CHMP. This may illustrate the difference between a compensatory model (MCDA) and a non compensatory one as BRR. In looking at the forest plot of relative risk (p. 41, Figure 4), it is tempting to look at the total lengths of the green and red bars: visually, red outweighs green, and this could lead to the conclusion that the risks outweigh the benefits. However, the longest red bar is for PML, but that criterion's modest weight in the MCDA led to its weighted effect approximately balancing only the PASI75 favourable effect, leaving a substantial benefit that was not overbalanced by the other unfavourable effects. In-the-head compensation of effects that do not take into account criterion weights, can lead to incorrect conclusions.

The definition of the Problem and of the Decision context proved to be rather difficult, because some of the options initially suggested to the 2009 assessors for their regulatory decision could not be addressed (lack of data which would have documented a Risk Minimisation measure). In addition, the balance to be assessed included short term Favourable Effects and long Tern Unfavourable one, as well as frequent mild effects versus rare serious irreversible one.

Similarly, the choice of the context for the modelling exercise didn't take into account some evolving competitive environment (ref 5), which probably played a role in the historical decision made in 2009. However, based on the framework initially chosen (placebo comparison) and the medical data which documented it, the MCDA model was relatively stable to significant changes in the assumptions made during the Decision Conference. This method could also easily accommodate the large heterogeneity of the relevant medical data as provided in a post-marketing decision context. In this respect, the BRAT framework proved to be more rigid in its application to non comparative data, and may be better suited in a Submission context with more homogenous data and units. BRAT would also require a quantitative method allowing for the integration of Benefit and Risk in a single score, should the decision-maker require or wish it.

Despite the rather large availability of efalizumab data in the public domain, the selection and extraction of data for subsequent representation in a proper Effect Table proved to be cumbersome and difficult. The same comment can be made for the set up of the Data Table within the BRAT framework, which suggests that the use of any model would require both appropriate biostatistical expertise and time/resources. This point would have to be taken into account if such B-R evaluation has to be made in an emergency or crisis situation.

For the methods requiring the building of a consensus (either on the criteria to be chosen, the measures applied to these criteria or the preference elicitation), the importance to be given to this "social process" is to be strongly emphasised and requires a face to face whole day meeting if not even more. In this context, the preference elicitation exercise should be well structured in order to obtain best results without biasing any option. In this discussion, it appears clearly that the perspective chosen (regulator, patient, Company) is of utmost importance as it impacts all of the subjective data used in the model, and to some extend the choice of the objective (medical) one.

The construction of a model intended to make a medical benefit-risk problem explicit, requires some assumptions and some choices (which criteria to select for a Value Tree, which data to select for an Effects table, which perspective to adopt etc.). Whilst it may be questioned on how much is lost in translating a complex model to simple one, the medical significance and justification of these choices should take precedence on any other one.

This also opens a possible discussion on those medical situations where a modelling is recommended given the complexity of a situation, as opposed to those where no modelling is necessary because the decision can be made quickly with an acceptable degree of consensus across all stakeholders.

# 8 References

1. HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/MEDICINES/HUMAN/MEDICINES/000542/HUMAN_MED_001012.JSP&MID=WC0B01AC058001D124

2. HAMMOND, JS AND AL, (1999) SMART CHOICES: A PRACTICAL GUIDE TO MAKING BETTER DECISIONS. HARVARD UNIVERSITY PRESS

3. PM COPLAN, RA NOEL, BS LEVITAN, J FERGUSSON AND F. MUSSEN, DEVELOPMENT OF A FRAMEWORK FOR ENHANCING THE TRANSPARENCY, REPRODUCIBILITY AND COMMUNICATION OF THE BENEFIT–RISK BALANCE OF MEDICINES, CLINICAL PHARMACOLOGY & THERAPEUTICS (2011) 89 2, 312–315. DOI:10.1038/CLPT.2010.291

4. PHILLIPS LD (2007). DECISION CONFERENCING, IN ADVANCES IN DECISION ANALYSIS: FROM FOUNDATIONS TO APPLICATIONS, W. EDWARDS, RF MILES AND D VON WINTERFELDT, EDS, CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE

5. BRITISH ASSOCIATION OF DERMATOLOGISTS GUIDELINES FOR USE OF BIOLOGICAL INTERVENTIONS IN PSORIASIS 2005 C.H. SMITH, AND AL, BRITISH JOURNAL OF DERMATOLOGY VOLUME 153, ISSUE 3, PAGES 486–497, SEPTEMBER 2005

# 9 Appendix

1. Please see “Supplement 1 to Wave 1 case study report Efalizumab” for the original documentations on PrOACT-URL, PhRMA BRAT and Efalizumab effects table. Available on http://www.imi-protect.eu/benefitsRep.shtml.

2. Please see “Supplement 2 to Wave 1 case study report Efalizumab” for the original report on the Decision Conferencing for the Multi-Criteria Decision Analysis. Available on http://www.imi-protect.eu/benefitsRep.shtml.