Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

IMI PROTECT Benefit-Risk Group

PATIENT AND PUBLIC INVOLVEMENT REPORT version 1.0

Recommendations for Patient and Public Involvement in the assessment of benefit and risk of medicines

Authors	Organisation
Kimberley Hockley [1] Deborah Ashby[2] Subhakanta Das Christine Hallgreen Shahrul Mt-Isa Ed Waddingham Richard Nicholas	Imperial College London
Susan Talbot	Amgen
Isabelle Stoeckert	Bayer
Georgy Genov	European Medicines Agency
Yasemin Dil Jo Groves Rebecca Johnson Alison Lightbourne Jeremiah Mwangi Rachel Seal-Jones	International Alliance of Patients' Organisations
Adam Elmachtoub	Massachusetts Institute of Technology
Craig Allen Andrew Thomson	Medicines and Healthcare products Regulatory Agency
Emanuel Lohrmann Alain Micaleff[2]	MerckSerono SA
Richard Nixon	Novartis
Joanne Treacy Lesley Wise	Takeda

[1] Team leader of the IMI PROTECT Work Package Five Patient and Public Involvement Workstream

[2] Co-leaders of IMI PROTECT Work Package Five

Disclaime: 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

Abstract

Background

Pharmacoepidemiological Research on Outcomes of Therapeutics in a European Consortium (PROTECT) is a collaborative project aiming to strengthen the monitoring of the benefit-risk of medicines in Europe. Work Package 5 within PROTECT investigated methods for integrating benefit and risk data with the perspectives of patients, physicians, regulators and other stakeholders (such as society as a whole) in order to make treatment decisions. One topic of particular interest within the group was the inclusion of wider societal viewpoints through Patient and Public Involvement (PPI). A team was established to investigate the concept of PPI and its potential application to regulatory benefit-risk decision-making and develop a resource for those working in this area.

Aims

- To assess where and when in the benefit-risk assessment decision-making roadmap PPI is desirable and/or feasible.
- To elicit weights from patients and the public regarding benefit and risk outcomes for the treatment of relapsing remitting multiple sclerosis (RRMS).
- To test visual formats for the communication of different stages of benefit-risk assessment with patients and the public and healthcare professionals.
- To create a summary of the recommendations and PPI work suitable for laypeople.

Methods

- Benefit-risk methods and frameworks were examined to assess how PPI could be employed in conjunction therewith
- The methods Swing-weighting, MACBETH, DCE, and AHP were used in parallel to elicit outcome preferences directly from multiple sclerosis patients in London
- Surveys were carried out to test laypeople's and healthcare professionals' understanding of visual representations of benefits and risks of weight los treatments
- A public website was established summarising the work package's research for both professionals and lay readers.

Interim results

- A framework for applying PPI in benefit-risk assessment was developed
- Due to ethical approval delays, other results are pending

Conclusion

There is greater clarity regarding the potential application of PPI in benefit-risk assessment, but we must await further results before commenting on how the methods fare in practice.

Contents

List of figures

List of tables

Abbreviations

Abbreviation	Description
ADR	Adverse drug reaction
AE	Adverse event
AHP	Analytic hierarchy process
CHMP	Committee for Medicinal Products for Human Use
DCE	Discrete choice experiment
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EU	European Union
FDA	Food and Drug Administration
FFD	Fractional factorial design
IAPO	International Alliance of Patients' Organisations
IMI	Innovative Medicines Initiative
MACBETH	Measuring Attractiveness by a Categorical Based Evaluation Technique
MAH	Marketing Authorisation Holder
MCDA	Multi-criteria decision analysis
MNL	Multinomial logit
MS	Multiple sclerosis
NHS	National Health Service
PML	Progressive multifocal leukoencephalopathy
PPI	Patient and public involvement
PROTECT	Pharmacoepidemiological Research on Outcomes of Therapeutics in a European Consortium
RRMS	Relapsing remitting multiple sclerosis
SAE	Serious adverse event
SPC	Summary of Product Characteristics
WP5	Work Package Five

1 Introduction

1.1 IMI PROTECT

Pharmacoepidemiological Research on Outcomes of Therapeutics in a European Consortium (PROTECT) is a project set up under the Innovative Medicines Initiative (IMI). Its goal is to strengthen the monitoring of the benefit-risk of medicines in Europe by developing a set of innovative tools and methods which will enhance the early detection and assessment of adverse drug reactions from different data sources, and enable the integration and presentation of data on benefits and risks. These methods will be tested in real-life situations in order to provide all stakeholders (patients, prescribers, public health authorities, regulators and pharmaceutical companies) with accurate and useful information supporting risk management and continuous benefit-risk assessment. PROTECT is a collaboration between 34 private and public sector partners and is coordinated by the European Medicines Agency (EMA). There are a total of six work packages in PROTECT, each with a different area of focus. The patient and public involvement (PPI) work described in this report is part of Work Package Five (WP5). The goal of WP5 is to develop methods for use in benefit-risk assessment, including an examination of the underpinning modelling, the presentation of results, and application of PPI.

1.2 Work Package Five and Patient and Public Involvement

From the outset of WP5 it was anticipated that, "the perspectives of patients, physicians, regulators and other stakeholders such as societal views" would be investigated in the context of benefit-risk assessment (WP5 Charter). Although there was commitment in the WP5 Charter to investigate PPI, it did not explicitly describe potential research activities which could be implemented to meet this aim; these were formulated later as the work package progressed and WP5 became more familiar with the field of PPI and benefit-risk assessment methodologies. A strong interest in patient and public involvement (PPI) from the PROTECT WP5 Wave 1 and Wave 2 case study task forces led many discussions concerning the application of PPI principles and patient and public values to the regulatory benefit-risk assessment of medicines. When applying PPI within a regulatory decision-making context, WP5 noted that there were a distinct lack of guidelines which address the specific methodological aspects and challenges which may arise in its practice. Initial PPI work to elicit patient and public preferences took place in the Wave 2 rimonabant and natalizumab case studies. However, PPI work in the natalizumab case study was not completed in Wave 2 as the elicitation activities required more time than originally anticipated. However, following this, WP5 still felt it was desirable to continue this work and dedicated a team to complete it and further investigate the concept of PPI and its potential application to regulatory benefit-risk decision-making in more depth.

1.3 Background

Before the team began, we felt that it was important to define the term PPI and how we would use it within this report.

1.3.1 Defining Patient and Public Involvement

Patient and public involvement relies on an active partnership between health professionals (e.g. regulators, physicians, the pharmaceutical industry, academics) and patients and the public. For the purposes of this report, the words "patient and public" and "involvement" were defined as:

Patient and public: Clinical trial participants, patients and potential patients, disabled people, parents and guardians, people who use health and/or social care services, carers, members of the public, and the organisations who represent the interests of these individuals.

Involvement: An active partnership between the patients, the public, and regulators in the decision-making process, rather than the use of patients and the public as the 'subjects' of decision-

making. Patient and public involvement can be defined as decision-making 'with' or 'by' patients and the public, rather than 'to', 'about' or 'for' the patients and the public.

These definitions were adapted from Involve (http://www.invo.org.uk/), a national advisory group which supports greater public involvement in National Health Service (NHS).

Although we chose to adopt these definitions in WP5, we would like to acknowledge that varying definitions of "patient and public" can be adopted in relation to the benefit-risk assessment of medicines. For example, some may refrain from including "members of the public" in their definition, as they may not be directly affected by the indication under consideration, and/or understand how it can affect quality of life. Therefore, the involvement of the general public not be desirable because their values and preferences might be considered less relevant and/or valid to the benefit-risk decision-making context than those directly affected.

The WP5 PPI team chose to adopt an inclusive definition of "patient and public" because our focus was to test "proof of concept" and demonstrate: (a) the feasibility of applying PPI to benefit-risk assessment methodologies which centred on eliciting the values and preferences, (b) which benefit-risk visualisations are most comprehendible/preferred/trustworthy, and (c) developing a section of the WP5 website for patients and the public.

1.3.2 The importance of involvement

The ultimate goal of healthcare policy decisions is to improve health and quality of life for patients. Since patients are arguably the best judges of the impact of these decisions on their lives, there is a strong moral/ethical argument that they should play a meaningful role in the decision making process. Engaging patients in health policy decision-making helps to ensure that policies reflect patient and caregiver needs, values, preferences and capabilities. Often methods of dealing with questions of benefit and risk of medicines do not include the perspective of the individual patient but patients and the public need to be involved in shaping the environment in which healthcare processes are designed in order to ensure a patient-centred approach.

There are many reasons why it is desirable to engage patients and members of the public in regulatory decision-making. Involve (http://www.involve.org.uk/) states that there are four main benefits of public involvement in policy processes: (1) improved governance; (2) social capital and social justice, (3) improved quality of services, projects and programmes; and (4) capacity building and learning (Involve, 2011). Specifically in the field of regulatory decision-making, regulators are increasingly expected to engage in PPI as it increases the legitimacy of the licensing process by improving transparency, and thus enhances public trust (Eichler et al., 2012). Despite this, there are many barriers to involvement including: a lack of awareness, difficulties identifying patient groups and patients, questions of representation, credibility and independence, lack of evidence of impact, lack of knowledge (on both sides) and understanding, and a lack of resources.

1.4 Research aim

The overarching aim of the PPI team was to develop a resource for those who wish to incorporate PPI into benefit-risk decision-making in the context of medicines. Our scope would specifically focus on four key areas: (a) patient and public involvement and the benefit-risk assessment roadmap, (b) preference elicitation from patients and the public regarding benefit and risk outcomes which includes the evaluation of different benefit-risk assessment methodologies; (c) visual representations; (d) Communicating the outputs of WP5 to participants, patients, and the public.

Each of these key areas had specific aims, shown in Table 1-1.

Table 1-1 Key areas of research and aims

Key area	Aim
Patient and public involvement and the benefit-risk assessment roadmap	To assess where and when in the benefit-risk assessment decision-making roadmap PPI is desirable and/or feasible.
Preference elicitation from patients and the public	To elicit weights from patients and the public regarding benefit and risk outcomes for the treatment of relapsing remitting multiple sclerosis (RRMS) using the benefit-risk methodologies of swing-weighting, MACBETH, DCE, and AHP (see Section 3.3.3).
Visual representations	To test visual formats for the communication of different stages of benefit-risk assessment with patients and the public and healthcare professionals.
Communicating the outputs of WP5 to participants, patients, and the public	To create a lay summary of the recommendations and PPI work.

This report describes the work embarked upon to address these aims. It is primarily derived for the purposes of PROTECT, but may also have wider application to other decision-making contexts where PPI is required.

1.5 Report structure

This report consists of four sections:

1. Introduction (this section)
2. Patient and public involvement and the benefit-risk assessment roadmap
3. Preference elicitation from patients and the public
4. Discussion

The work done on visual representations and communicating WP5's outputs will be reported in due course. Preliminary outputs can be seen on the PROTECT BR website at http://www.protectbenefitrisk.eu/PPI.html .

Following this, there is an additional section where overall conclusions are presented and discussed.

2 Patient and Public Involvement in systematic methods of benefit-risk assessment

2.1 Introduction

Depending on the aim and objectives of PPI when applied to benefit-risk methodologies, there are multiple stages in the product lifecycle where involvement can occur; varying degrees to which patients and the public can be involved; and multiple stages in each benefit-risk methodology where involvement can be applied.

2.2 Methods

Via a careful examination of the descriptive and quantitative benefit-risk methodology frameworks used within PROTECT WP5—which make explicit recommended steps to guide benefit-risk assessment, it is possible to carefully evaluate which specific points in the decision-making process are most suitable for the application of PPI.

The WP5 Recommendations team analysed the various benefit-risk methodologies which were applied to the case studies in order to determine the decision-making roadmap. Each stage in the roadmap represents a point in the decision-making process where some or all decision-maker(s) are required to come together in order to evaluate current findings, and make interim decisions. The team identified five key stages: planning, evidence gathering and preparation, analysis, exploration and decision and dissemination.

In this section, we provide a summary of the key steps in the critical pathway common to both PrOACT-URL (http://www.protectbenefitrisk.eu/PrOACT-URL.html) and BRAT (http://www.protectbenefitrisk.eu/BRAT.html): (1) planning, (2) evidence gathering and data preparation, (3) analysis, (4) exploration, and (5) conclusion and dissemination, and discuss how we believe PPI can be applied to each of the activities performed within each stage.

2.3 Results

In this section we describe how PPI can be applied to different stages of the benefit-risk assessment process (Section 2.3.1), with different levels of involvement (Section 2.3.2), and at varying time points during the lifecycle of a product (Section 2.3.3). We also discuss how current proposals for adopting PPI in systematic methods of benefit-risk assessment (Section 2.3.4) and the application of PPI throughout the entirety of the benefit-risk decision-making process (Section 2.3.5). Finally, we discuss where we believe where and when in the benefit-risk assessment decision-making roadmap PPI is desirable and/or feasible (Section 2.3.6).

2.3.1 Varying stages

Although PPI may occur throughout the decision-making process, there may be specific stages in each benefit-risk methodology where PPI may be more feasible or desirable. For example, PPI may be considered important during (a) the selection, inclusion and exclusion of relevant outcome measures, or (b) the ranking and weighting of outcome measures. It is important to note that these two examples are not exhaustive, and there are many more examples of where patient involvement may be considered important.

2.3.2 Varying degrees

It is necessary to establish the desired level of involvement and consider how much of an active role patients and the public should take in the decision-making process. The degree of involvement may vary according to which decision-making stage it is applied to in a specific methodology, or it may remain at a constant level throughout the whole process (i.e. from setting the decision context to final reporting). There are two main categories which reflect the degree of patient involvement in regulatory decision-making (adapted from Hanley et al. (2001)):

Consultation: regulators elicit the patient perspective to inform the decision-making stage or entire decision-making process.

Collaboration: regulators and patients and the public form an active partnership and jointly participate in the decision-making stage or entire decision-making process.

2.3.3 Varying time points in the product lifecycle

It is also necessary to acknowledge that PPI may occur at any time point in the product lifecycle. For example, PPI applied to the design of early stage clinical trials might affect which outcomes are investigated in the Development Programme, which in turn will directly affect which outcomes have sufficient data to be included when assessing the benefit-risk balance for marketing authorisation decisions. Another potential application for PPI is when choosing the dose regimen following a dose ranging clinical trial. One of the aims of such studies is to identify the regimens in the therapeutic window where sufficient efficacy is demonstrated whilst maintaining acceptable safety. The choice is essentially to assess the regimen that gives the optimal benefit-risk balance. In WP5 the case studies specifically focus on the decision-making time point of evaluating the benefit-risk balance during initial marketing authorisation decisions, and/or when it was necessary to re-evaluate the benefit-risk balance following the occurrence of adverse events detected during post-marketing surveillance. In this section we focus on these two specific time points, and examine where PPI can potentially occur in benefit-risk methodologies rather than focussing on stages of the product lifecycle to which we can apply PPI.

2.3.4 Current adoption of PPI in the benefit-risk assessment process

PPI is currently recommended by PrOACT-URL and BRAT frameworks for the stages of preparation and analysis (http://www.protectbenefitrisk.eu/recommendations.html).

The involvement of patient advocates is indicated several times in "The PhRMA BRAT Framework for Benefit-Risk Assessment: User's Guide to the Process" (The Benefit-Risk Action Team, 2011). It specifically suggests that patient advocates may be included as optional stakeholders in the BRAT framework development team and provide external input at two key stages: Step 2: Identify Outcomes, and Step 5: Assess Outcome Importance. Additionally, PrOACT-URL briefly mentions the role of patients in a few steps, although does not detail their explicit involvement.

While the brevity of PPI instructions for adoption in both frameworks is noted, it should be acknowledged that both frameworks are intended to guide the decision-making process rather than consist of a fixed set of exhaustive and prescriptive tasks to complete. Therefore one would not expect either framework to fully set out where PPI could be applied. However, their description of decision-making processes can still provide insight by helping to identify where PPI can potentially be adopted.

2.3.5 Potential adoption of PPI throughout the entirety of the benefit-risk decision-making process

Many believe that PPI can and should be applied throughout the lifecycle of a product. This includes incorporating PPI into all the regulatory benefit-risk decision-making processes that determine the marketing authorisation decision.

Although there are strong arguments to support the application of PPI throughout the entirety of the benefit-risk decision-making process, it should be acknowledged that this might not be possible due to the lack of methodological guidelines, and/or a lack of time and resources or other practical restrictions. As PPI is still a relatively novel concept in benefit-risk decision-making, many decision-makers do not know where it can be applied, what might be necessary or helpful for its application, what potential barriers may exist, and how involvement can be most meaningful and valid.

Because of these reasons, it is very difficult to try to simultaneously apply PPI to multiple points throughout the entirety of the marketing authorisation assessment and post-marketing surveillance. Therefore we consider that the best approach is to systematically approach each step of the pathway in turn; and investigate what would be the most meaningful and beneficial way to involve patients at each step.

2.3.6 The benefit-risk assessment roadmap and application of PPI

In this section we provide a summary of the key steps described in the benefit-risk assessment roadmap: (1) planning, (2) evidence gathering and data preparation, (3) analysis, (4) exploration, and (5) conclusion and dissemination, and discuss how we believe PPI can be applied to each of the activities performed within each stage.

2.3.6.1 Planning

In the planning stage the problem statement is established, and a number of key contextual factors are made explicit and documented. They include critical issues such as stating who the decision-maker is, the perspective they are adopting (e.g. if they are deciding for themselves or on behalf of others), the indication that the product is aiming to or has received a marketing authorisation for and severity of the condition and the severity of the condition, the target population, and the treatments under consideration in the decision-making scenario. The relevant favourable effects and unfavourable effects to be included in subsequent steps of the decision-making process are also identified based on expert opinion, literature searches and relevance to key stakeholders. It is also decided which benefit-risk methodologies will be used—e.g. BRAT or PrOACT-URL.

All stakeholders—including patients and the public, can be included when defining the decision problem. Patients and the public are key stakeholders and can potentially be included in the group of decision-makers, and/or decision-makers may elicit patient and public views and opinions to incorporate into the decision-making process. If the patient and public perspective is insufficiently addressed and/or incorporated into the problem statement in the planning stage, the decision-making scenario may inadequately represent the immediate concerns of those who have first-hand knowledge and experience of the indication under consideration. An incorrect or insufficient problem statement which does not appropriately reflect patient and public concerns can potentially lead a significant amount of resources spent on benefit-risk assessments where the results may be of questionable benefit to patients. One important difference between regulators and patients which needs to be addressed, is that regulators make decisions at a population level, whereas patients make decisions at an individual level. It is necessary for future work to acknowledge and account for this fundamental difference when defining the problem statement.

Hypothetically, everyone directly affected by the decision should participate in decision-making discussions. There may be differences in opinion regarding which critical issues should be included in the decision problem, e.g. the objectives of the assessment, options under consideration, and facts to be included in the analysis. Ideally, there should be agreement between all stakeholders. Many consider it desirable to organise meetings which bring stakeholder groups together for clearly structured discussions to establish the purpose and context of decision-making, and come to consensus. There are a multitude of different skillsets, knowledge bases, and expertise represented by each of the stakeholder groups, e.g. empirical, statistical, decision analysis, medical, regulatory, scientific and epidemiological. All of these perspectives can be complementary. It is not currently clear what the best way of balancing different or opposing perspectives may be in benefit-risk decision-making.

Patients and the public are likely to lack the technical expertise to identify and select which benefit-risk assessment methodologies are best suited to address a specific problem statement, and PPI for this task may be limited. In this step, professional expertise may be necessary to lead the decision-making process. Additionally, patients and the public are not likely to know how their values and preferences can be interpreted and incorporated into the regulatory assessment of medicines. Again, professional expertise may be necessary in the acquisition,

interpretation, and incorporation of patient preference values into the regulatory assessment of a medicine. However, PPI can potentially be possible through presenting patients and the public with options of how they can be involved in the decision-making process, and they can be given the option to select where they feel their input would be most valuable. If PPI is to be organised by experts, clear communication of the activities and processes to be undertaken as part of the decision-making methodologies is essential.

Patients and the public can provide justification on which favourable and unfavourable effects they believe should be included in the benefit-risk assessment. The only disadvantage is that patients and the public might consider it desirable to include favourable and unfavourable effects for which data has not been collected or has been published. There are a multitude of sources which can be used to identify benefits and risks for inclusion—both epidemiological and empirical; the patient perspective should be included. Patients have a wealth of understanding and expertise in how a specific indication can affect quality of life and activities of daily living, and even prior experience of other medications used to treat the same indication. It is imperative that patient and public knowledge and expertise is not disregarded; decision-makers should aim to ensure that all sources of evidence appropriate to the decision problem are accounted for—including patient knowledge, while also accounting for other forms of objective scientific data.

It is important to document the process and results of decision-making methodologies in a transparent manner, ensuring that they are available to all stakeholder groups. Clear, user friendly and accessible materials should be developed to report not only the planning stage, but all of the decision-making processes. It needs to be determined and agreed where, when, and how the processes and results of the methodologies are communicated to patients and the public in suitable format(s), with the use of visualisation(s) where appropriate. Once the process of a decision-making methodology becomes transparent, it automatically opens up the process to audit; it becomes clear where the patient perspective has been adopted. However, this also means that if the patient and public perspective is investigated but not explicitly incorporated into the results, they may feel disenfranchised which could result lack of trust and motivation to participate in the future.

2.3.6.2 Evidence gathering and data preparation

The evidence gathering and data preparation stage is when evidence related to the benefit-risk assessment is identified and extracted, including data on the favourable and unfavourable effects associated with a given treatment. The source data is clearly documented, together with details of any data manipulations.

A great amount of technical expertise is required to gather evidence and prepare data for a benefit-risk assessment. Epidemiological and statistical expertise is required to analyse and interpret data, and clinical and patient and public expertise can be used to ensure that appropriate judgements are made regarding the relevance of favourable and unfavourable effects criteria. Although patients and the public might lack technical expertise (e.g. data transformation, aggregating evidence from multiple sources and exploring statistical uncertainty), it is imperative that patient and public knowledge and expertise is not disregarded; decision-makers should aim to ensure that all sources of evidence appropriate to the decision problem are accounted for—including patient knowledge, while also accounting for other forms of objective scientific data.

Despite this, it should be noted that when extracting favourable and unfavourable effects data on the performance of each treatment under consideration, the information which patients and the public may know might be anecdotal and subject to bias. There is also a distinction between publicly available information, and confidential information which might be listed in the submission dossier—only accessible to regulators and the pharmaceutical industry.

The evidence gathering and data preparation stage should be clearly documented for replication purposes. It may also be useful to document the process for other stakeholder groups; communication can enhance transparency and

trust. However, there are many challenges when trying to articulate evidence gathering and data preparation, which may include complex data manipulations, to a lay audience. There may be a reduction in the level of reported detail, and whilst a reduction in the number of technical terms used may increase comprehension, it may also result in a loss of precision.

2.3.6.3 Analysis

The analysis stage is where performance data are analysed to quantify the magnitudes of benefits and risks for the treatment(s) of interest. Depending on the purpose and context of the benefit-risk assessment, the favourable and unfavourable effects may be weighted and integrated to provide a quantitative measure of the benefit-risk balance. Descriptive (qualitative) or quantitative benefit-risk methodologies may be applied according to the complexity of favourable and unfavourable effects data and/or level of transparency desired by the decision-maker(s). Quantitative decision models such as MCDA disaggregate a complex problem into simpler components that are easier to understand and weigh up, and then use methods to integrate the components into a measure of the overall benefit-risk balance.

It may be possible for a decision at the analysis stage to be made implicitly, based on the evidence from the previous stage. If the decision is made qualitatively or implicitly, this indicates that it was not necessary for the decision-maker(s) to quantify the difference in importance between the favourable and unfavourable effects in order to make a sound decision. However, the patient and public perspective could be sought to verify if they would agree with the decision-maker(s) conclusion. If there is a disagreement when a decision is made implicitly, and/or a greater degree of transparency is required to articulate the thought processes and interim decisions of the decision-maker(s), quantitative decision decision-making methodologies may help.

Quantitative decision-making methodologies have the advantage of identifying where discrepancies may lie when comparing stakeholder perspectives, and see if they can be resolved. There may be variations in how different stakeholders evaluate the importance of favourable and unfavourable effects, and quantitative models lend themselves very readily to the application of PPI. Once patients and the public understand the favourable and unfavourable effects included in the model, it is possible to elicit their weights and then integrate the components into a measure of the overall benefit-risk balance. However, quantitative models demand more technical expertise and resources than descriptive methods.

2.3.6.4 Exploration

In the exploration stage, the results are assessed for robustness and sensitivity to the various assumptions and sources of uncertainties. Areas of uncertainty which should be investigated in quantitative benefit-risk methodologies include data on the favourable and unfavourable effects, and preference weights, and it is essential to establish how robust the results are while accounting for these uncertainties. It is also important to explore further the consequences of a benefit-risk decision, and consider whether the results of the benefit-risk assessment may inform related decisions on risk management plans or benefit-risk assessments of similar treatments.

The level of PPI is likely to be minimal in assessing the robustness and sensitivity to the various assumptions and sources of uncertainties. This is because of the high level of technical and statistical expertise required. Despite this, if preference weights were elicited in the analysis stage with PPI, these are now further examined to determine the robustness of the final decision regarding the benefit-risk balance. It may be possible to use elicitation methods to to establish plausible weight ranges for exploration purposes. Assessors can then go on to look at how variations in the inputs of a quantitative model—in this case elicited preference weights, can affect the resulting benefit-risk balance. These variations may exist due to reasons such as poorly framed questions, the generalizability of participant weights compared to patient and public populations, and sampling errors. Attitudes towards uncertainty may vary between stakeholders, depending upon the degree of risk averseness or desired level of caution.

2.3.6.5 Decision and dissemination

The last stage is where the final decision is made and the results are communicated to relevant stakeholders. It makes explicit that the findings of the benefit-risk assessment have logically led to a decision that may influence future actions. It emphasises the need for a transparent audit trail of the whole assessment process from the planning stage to the exploration stage. This last stage of the process brings everything together and sets the stage for action to be taken.

Quantitative models often integrate multi-faceted problems to a single number representing the benefit-risk balance, and this number firmly depends on the preference values used or elicited by the model which may vary from one stakeholder group to another, and even within each stakeholder group. Stakeholders (including patients and the public) might feel concerned about the mathematical processes applied to their wealth of knowledge and experience and how this is combined with favourable and unfavourable data, either from trials at the point of marketing authorisation or from various sources after the point of marketing authorisation, such as observational studies, spontaneous reports or further trials. There may be a danger that the output of quantitative models may be misinterpreted, and without effective communication there may be a degree of distrust. Once again, it must be reiterated that the process and results of the decision-making process should be clearly and effectively communicated to all stakeholders in an appropriate format with unambiguous language.

2.4 Discussion

The aim of this section was to examine where and when in the benefit-risk assessment process PPI can occur. To meet this aim, the roadmap produced by the WP5 recommendations team was described and the compatibility of applying PPI to the tasks involved in each step was carefully evaluated.

By carefully examining the decision-making pathway defined by the benefit-risk assessment methodologies, the task of critically evaluating where PPI may be desirable and/or be feasible in the benefit-risk assessment process was made easier. PPI is a concept which is relevant throughout the decision-making process and can be applied to almost every part of the benefit-risk assessment process. Therefore, before deciding how to adopt PPI into a specific benefit-risk assessment scenario, it is essential for the decision-maker(s) to carefully consider what the aim and objectives of applying PPI are, and which resources are available.

One limitation to the work presented in this section is that a purely theoretical application of PPI is discussed and real world feasibility is not evaluated. Additionally, it is important to note that our critical evaluation of the roadmap was based on subjective opinions of where PPI can occur and other stakeholders (including patients and the public) might have different opinions of where they believe PPI can potentially occur, and where they feel it would be most desirable to apply the principles.

3 Preference elicitation from patients and the public: planning a case study involving people living with multiple sclerosis in a London NHS Trust

3.1 Introduction

The work in this section was planned as a result of the importance of patient and public involvement and increasing questions and challenges about the use of systematic methods of benefit-risk assessment. When regulators have to make a decision about the balance of benefits and risks, these decisions are often complex. The rationale behind why the final decision was positive or negative may be challenging to communicate. Systematic methods of looking at this balance may help with this and provide step by step instructions on how to approach decision-making and understand how patients and public value the benefits and risks of medicines through preference elicitation. Although there is strong justification to elicit preferences from patients and the public, there are two main challenges: (1) there are many different types of methods which can be used to elicit preferences, (2) they have not been formally tested with patients and the public. We planned to explore and address these concerns with the study described in this section which evaluates patient preferences for some of the favourable and unfavourable outcomes of natalizumab used to treat relapsing remitting multiple sclerosis.

3.1.1 Relapsing remitting multiple sclerosis

There are three main types of MS: relapsing remitting (which represents approximately 85% of MS cases) which can evolve into the second type, secondary progressive (characterised by disease progression and incomplete recovery following relapses), and the third type is primary progressive (which represents approximately 10 to 15% of MS cases) (MS Society, 2013). RRMS is characterised by attacks (relapses) between periods of no symptoms (remissions). A relapse is defined as "the appearance of new symptoms, or the return of old symptoms, for a period of 24 hours or more – in the absence of a change in core body temperature or infection" (MS Society, 2013). Symptoms can be physical (e.g. vision, balance, speech, tremor, bowel functioning) and/or can affect memory, thinking and emotions. Stress, infection, vaccination and pregnancy have been found by some to trigger relapses.

Relapses can range from mild to severe. If the relapse is mild, the individual may be treated at home. If the relapse is more severe, hospital treatment may be required. During a relapse, demyelination occurs; inflammation caused by T cells stimulates other immune cells and soluble factors, e.g. cytokines and antibodies to produce leaks in the blood–brain barrier. This in turn causes a number of other damaging effects such as swelling, the activation of macrophages, and the activation of cytokines and other destructive proteins (Compston and Coles, 2008). The duration of a relapse may range from a few days to many months, and they typically last for approximately four to six weeks. Following this, there is a period of remission, where individuals recover from the symptoms. However, if there is severe damage to the myelin, some symptoms may remain.

3.1.2 Indication of natalizumab

Natalizumab (Tysabri®), received marketing authorisation from the US Food and Drug Administration (FDA) from November 2004 to February 2005 for the indication of relapsing remitting multiple sclerosis (RRMS). The marketing authorisation was subsequently suspended due to the occurrence of PML (similarly to efalizumab), but was later reintroduced to the market with strict risk minimisation measures in June 2006—demand from patient organisations was a major factor which lead to its reintroduction. In the EU, the EMA granted marketing authorisation for natalizumab in June 2006. In 2009, the benefit-risk balance of the treatment was reassessed by the CHMP due to new reported cases of PML; the marketing authorisation was maintained with risk minimisation measures in place.

Natalizumab is a recombinant humanised monoclonal antibody treatment. It recognises and attaches to α4β1 integrin, found on the surface of most leucocytes, i.e. the white cells in the blood which are involved in the

inflammation process. This prevents leucocytes from travelling in the blood to the brain, and thus reduces the inflammation and nerve damage caused by MS (EMA, 2010).

Natalizumab is indicated as single disease modifying therapy in highly active RRMS for (EMA, 2009g):

a) "Adult patients aged 18 years and over with high disease activity despite treatment with a beta-interferon, i.e. at least 1 relapse in the previous year while on therapy, and have at least 9 T2-hyperintense lesions in cranial Magnetic Resonance Image (MRI) or at least 1 Gadolinium-enhancing lesion or an unchanged or increased relapse rate or on-going severe relapses, as compared to the previous year"; or,
b) "Adult patients aged 18 years and over with rapidly evolving severe relapsing remitting multiple sclerosis defined by 2 or more disabling relapses in one year, and with 1 or more Gadolinium enhancing lesions on brain MRI or a significant increase in T2 lesion load as compared to a previous recent MRI."

300mg of natalizumab is administered by intravenous infusion to RRMS patients once every four weeks in a hospital or clinic. The duration of infusion is one hour, and the patient must be monitored for an additional hour afterwards. Treatment is reconsidered in those who do not experience therapeutic benefit beyond six months, and there is a reassessment of the potential for benefit and risk after two years since initiation of the treatment.

3.1.3 Marketing authorisation history

Natalizumab received marketing authorisation in June 2006 from the EMA and is still authorised at time of writing. The full procedural history is publicly available and can be found on the EMA website (EMA, 2009g, EMA, 2010). Important events relating to authorisation decisions and PML are summarised from the online documents below:

- **November 2004**: FDA granted marketing authorisation for natalizumab in the US.
- **February 2005**: FDA suspended marketing authorisation in the US due to the occurrence of two cases of PML (one of which was fatal) reported in RRMS patients in the SENTINEL clinical trial: both patients had been treated with a combination of natalizumab and beta-interferon for more than two years. Later, a third case of PML was discovered on re-evaluation of the Crohn's Disease safety database for a patient enrolled in the ENACT clinical trial.
- **June 2006**: FDA reintroduced natalizumab to the market with strict risk minimisation measures—the demand from patient organisations was a strong factor leading to its reintroduction.
- **27 June 2006**: CHMP issued a positive opinion for granting marketing authorisation with special warnings in the product information and extensive risk minimisation measures, including physician information and management guidelines.
- **October 2008**: Two cases of PML were confirmed in RRMS patients following EMA authorisation; for both cases natalizumab was used as a monotherapy and the treatment was administered for approximately 17 and 14 months. In total 38,700 patients had been treated worldwide since approval, and 4,650 patients had received natalizumab in clinical trials.
- **July 2008 to October 2009**: 23 confirmed cases of PML were reported worldwide in patients with MS receiving natalizumab between July 2008 and October 2009, resulting in four deaths. By 20 January 2010, the total number of confirmed PML cases had risen to 31 worldwide, of whom 23 had received natalizumab for more than two years. The EMA stated that this is equivalent to around one case of PML for every 1,000 patients treated with natalizumab for two years or more.
- **22 October 2009**: The CHMP requested a review of the benefits and risks to decide if marketing authorisation for the product should be maintained, varied, suspended or withdrawn. This was followed by the occurrence of new cases of PML, and in consideration of the occurrence of Immune Reconstitution Inflammatory Syndrome (IRIS) in PML patients once natalizumab had been stopped and plasma exchange

(PLEX) and/or immunoabsorption had been implemented. The CHMP concluded that the benefit-risk profile of natalizumab remained positive.

- **November 2010**: The MAH analysed the data of 52 natalizumab treated patients with confirmed PML in respect of an association with prior immunosuppressant (IS) use. The presented data indicated that prior IS use increases the risk of PML independent of the duration of natalizumab therapy.
- **April 2011**: The risk of PML with natalizumab was reported to be higher if the patients (a) have received natalizumab for more than 2 years, and/or (b) have received prior immunosuppressant therapy, and/or (c) are anti-JCV antibody positive. The EMA stated that patients who have all three risk factors for PML have the highest risk of PML (approximately 9 in 1,000 patients treated).
- **April 2011**: CHMP stated that it was their belief that the benefit-risk balance remained positive for the MS population with less than 24 months exposure to rimonabant, even in the presence of prior immunosuppressive treatment and seroprevalence of JCV. However, they stated that post-marketing data does not provide complete evidence of how the risk factors could combine and affect the benefit-risk balance in long-term exposed patients. Therefore, the CHMP offered the opinion that one additional five year renewal on the basis of pharmacovigilance grounds is required.
- **April 2012**: Based on continuing evaluation of PML, the PML incidence rates were updated based on post-marketing data, resulting in changes to the PML incidence figures for patients with antibodies against JCV or more additional risk factors. The CHMP considered that these revised PML incidence rates are not significantly different from the numbers that were included in the previous version and their opinion on the benefit-risk balance did not alter.

3.2 Aim and objectives

In PROTECT WP5, PPI was considered to be most desirable and valuable in the weighting stage of each of the benefit-risk methodologies identified for testing. This is the stage where the benefits and risks of treatments—commonly efficacy and safety measures collected during clinical trials and post-marketing surveillance, are ranked and weighted. We considered it to be an important stage to investigate because the weights allocated to the favourable and unfavourable effects of treatments have the potential to substantially vary according to whose perspective is adopted, which consequently may have a large impact when determining the final benefit-risk balance.

The primary aim of this work was to test the feasibility of obtaining patient and public preferences on the benefits and risks of RRMS treatments using different methods of elicitation proposed for use in benefit-risk assessment. This was achieved through a feasibility study, where our objectives were:

- To test how the benefit-risk assessment methodologies of AHP, DCE, swing-weighting and MACBETH can be used to elicit preferences from patients and the public affected by RRMS (i.e. what is the feasibility of each method and how do the methods compare against one another)
- To elicit preferences and examine how consistent or inconsistent the elicited preferences are when compared across methodologies
- To examine probabilistic modelling in treatment performance and elicited weights
- To report, evaluate and compare the process of involvement in benefit-risk methodologies from an administrative and participant perspective

3.2.1 Involvement of patients and the public in this research

Ideally we desired meaningful and in-depth patient and public involvement which represented a range of stakeholders throughout the research described in this section. However, due to time and resource limitations this was not possible. In lieu of this, representatives from the International Alliance of Patients' Organizations who are

partners in IMI PROTECT Work Package Five provided input and feedback on behalf of patients and the public. As a member of the Patient and Public Involvement team, IAPO were involved in the design, management, and undertaking of the research and dissemination of results. They have actively worked on the protocol and have also provided guidance and advice for specific topics, e.g. methods of communicating to patients and reimbursement for participation. In addition to this, the online questionnaire has been piloted with individuals with multiple sclerosis for feedback (e.g. to assess comprehension and ease of completing tasks).

3.3 Methods

Here we outline the general approach we adopted when developing our study to elicit preferences from patients and the public. A brief description of each of the steps is provided in this section:

1. Planning and data preparation
2. Selecting which method(s) of eliciting preferences to use
3. Research procedure
4. Evaluating the process of preference elicitation

Detailed descriptions which relate to how each of the steps was applied are provided in the following sections of this report.

3.3.1 Planning and data preparation

Planning and data preparation included a number of tasks: specification of the decision problem (Section 3.3.1.1), development of the value tree (Section 3.3.1.2), and the extraction and synthesis of data (Section 3.3.1.3).

3.3.1.1 Specification of the decision problem

While planning our benefit-risk assessment, the problem statement had to be established and pre-emptively we needed to consider, make explicit and document a number of key contextual factors. We considered critical issues such: the decision problem, whose perspectives are being adopted, comparators, favourable and unfavourable outcomes, sources of evidence, time horizon and resource allocation.

3.3.1.2 Development of the value tree

The benefit-risk assessment of a medicine using formal methodologies (e.g. PrOACT-URL, BRAT, MCDA, SMAA) involves evaluating how multiple objectives are met. These objectives are frequently conveyed within a value tree. Ideally, the value tree should visually represent the favourable and unfavourable outcomes of interest and fulfil a pre-specified set of criteria:

- Completeness: The objectives included in the value tree are relevant to the decision-making problem and comprehensively define the degree to which the overall objective can be achieved.
- Operationality: Objectives are clearly defined, meaningful and assessable/measurable.
- Preference independence: The weight of one objective should remain unaffected by the performance of any other objectives.
- Avoidance of double-counting: Objectives should not overlap or act as surrogates for one another.

3.3.1.3 Data

We identified source documents, and information on the favourable and unfavourable effects of treatment for benefit-risk assessment; usually, this is efficacy and safety data collected from clinical trials and post-marketing surveillance. Ideally, we aimed to consider all available information from pre-specified data sources, which may have needed to be synthesised in order to provide accurate estimates for assessment.

3.3.1.3.1 Data extraction

This study intended to replicate a real life decision-making scenario. Our data sources were: European Public Assessment Reports, Scientific Discussions, Important Safety Information, Summary of Product Characteristics, Prescribing Information and academic publications relating to key clinical studies.

We extracted efficacy and safety data from these documents, assuming that they presented information on the most important data to consider for each of the case study treatments from a regulatory, industry, clinician and patient perspective. We were limited to publicly available sources—similarly to all of the other WP5 case studies, because it would provide our example with a high degree of transparency, allowing others to closely follow, replicate or compare different methodologies with our example.

3.3.1.3.2 Data synthesis

We collated publicly available data from the sources and statistical modelling will be employed to allow treatments to be compared using indirect evidence and to estimate the uncertainty of the data. For further information see section 3.3.7.

3.3.2 Research procedure

3.3.2.1 Participant selection and recruitment

3.3.2.1.1 Screening and enrolment

We planned to recruit a sample of patients receiving treatments for relapsing remitting multiple sclerosis from Charing Cross Hospital and St. Mary's Hospital, London, England. Treating clinicians screened potential participants during routine appointments. Those who met the inclusion criteria were provided with a study pack (a stamped addressed envelope containing an information sheet (Appendix 5.1), consent form for a focus group) (Appendix 5.2) and questionnaires (Supplement PPI 1 – Supplement PPI 5).

Patients took the study pack home where they could decide whether they would like to participate and complete the questionnaire. Participants who wish to attend a focus group will be asked to provide contact details. A stamped addressed envelope was included in the study pack for returning the questionnaire. A subset of participants who expressed an interest in participating to the focus group were then invited to a decision conference. The right of the individual to refuse participation was respected and they did not have to provide reason or justification. Once participants enter the study, they were free to withdraw at any time without giving reasons and without prejudicing further treatment.

Ethical approval was received for the study through IRAS (#144663).

3.3.2.1.2 Inclusion and exclusion criteria

Potential participants were included in our study if they were:

- Aged 18 years old or above
- Have relapsing remitting multiple sclerosis
- Receiving treatment the MS Clinic at Charing Cross Hospital, Hammersmith or the Multiple Sclerosis Clinic at St. Mary's Hospital, Paddington

Potential participants were excluded if they were:

- Not fluent in spoken and written English
- Cognitively and/or visually impaired
- Clinician discretion (e.g. patient appears distressed, angry, or upset)

3.3.2.1.3 Participant information sheet

The participant information sheet is presented in the Appendix 5.1.

3.3.2.1.4 Consent

A completed, returned questionnaire was judged as implied consent. For the decision conferences, written consent was required. An information sheet and consent were provided in the study pack for participants to review and consider.

3.3.3 Selecting which systematic methods of benefit-risk assessment to use

We chose to implement the preference elicitation methodologies of AHP, DCE, swing-weighting and MACBETH. The reason we chose these methodologies is because they are frequently used and we wanted to perform a critical analysis of how they could be implemented from a PPI perspective. Here we describe the general steps of each of the benefit-risk assessment methodologies applied in this study. Section 3.5.6 (analytic hierarchy process), Section 3.5.7 (discrete choice experiments), Section 3.5.4 (swing-weighting) and Section 3.5.5 (MACBETH) will contain additional details about how these methods were implemented.

3.3.3.1 Analytic Hierarchy Process

AHP is another decision-making method that can take into account multiple risks and benefits simultaneously. There are several steps involved in the design and analysis of an AHP which overlap with other decision-making methods such as MACBETH and MCDA.

(1) Defining criteria and sub-criteria

The first step is to define a set of criteria and sub-criteria necessary for decision-making. The number of criteria and sub-criteria should be limited and cover the most important ones involved in the decision-making process.

(2) Define alternatives

Potential solutions to the decision-making problem are determined. The performance of each solution among criteria or sub-criteria is then defined.

(3) Criteria and sub-criteria weighting

The weighting of criteria and sub-criteria is then carried out. Weighting is performed in a similar fashion to MACBETH, with the exception that consistency is not checked. AHP uses a quantitative scale with a range of 1 to 9 (Table 3-1). The weighting process takes place as follows: the respondent is first asked which of two criteria is most important, and second to quantify the intensity at which it is more important. This is performed for each pair of criteria, and each pair of sub-criteria within each criterion. This means that for an AHP with three criteria, each with three sub-criteria (Figure 1), a total of 12 (i.e. 3 + 3 x 3) pairwise comparisons need to be made.

Table 3-1 Analytical hierarchical process (AHP) weighting scale

Intensity of importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgement moderately favour one element over another
5	Strong importance	Experience and judgement strongly favour one element over another
7	Very strong importance	One element is favoured very strongly over another; its dominance is demonstrated in practice

Intensity of importance	Definition	Explanation
9	Extreme importance	The evidence favouring one element over another is of the highest possible order of affirmation
Intensities of 2, 4, 6, and 8 can be used to express intermediate values. Intensities of 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance.		

(Saaty, 1990)

(4) Weighting of alternatives

The next step is to weight the performance of alternatives at the lowest levels of AHP hierarchy. Similarly to weighting criteria, pairwise comparisons of alternatives are made according to their performance. For an AHP with 3 criteria, each with 3 sub-criteria, and 3 alternatives (Figure 1), this means that a total of 27 (i.e. 3x3x3) comparisons need to be made. In some circumstances where it is justifiable, the weighting of alternatives can be skipped for a subset of comparisons. In these cases, assumptions need to be made on how the weight of each alternative is calculated, such as by assuming a linear relation between weight and performance.

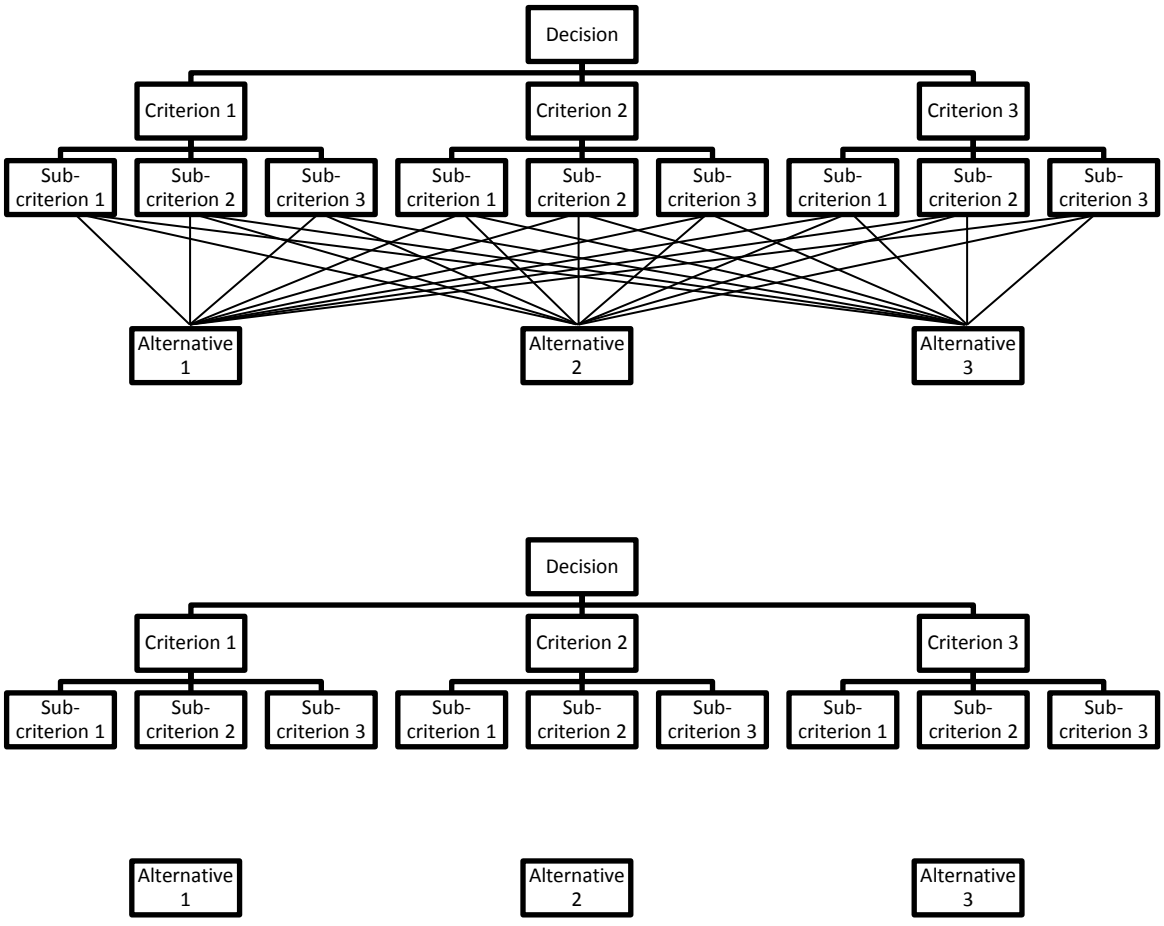


Figure 1 Illustration of an analytical hierarchical process including 3 criteria, each with 3 sub-criteria, and 3 alternatives

(5) Calculation of weights and the consistency ratio

Once the weighting process is completed, the weights of criteria, sub-criteria, and alternatives can be calculated. This is normally achieved through matrix algebra, where the most suitable combination of weights which fit the comparisons is determined. As consistency is not guaranteed or strictly required in AHP, a consistency ratio can also be calculated, for which a value of 0.1 or less is generally considered as suitable.

(6) Interpretation

The ranking of alternatives is made by descending weight, and the alternative with the highest weight can be considered as the preferred option. However, results should be interpreted with care if the consistency ratio is higher than the threshold of 0.1.

3.3.3.2 DCE

Discrete choice experiments (DCE) are surveys which observe choices and measure preferences. They provide explicit measures of benefit and risk valuation for assessing alternative treatment options, by evaluating the choice behaviour of participants to infer values.

Four steps are used to guide the application of DCE:

1. Identify attributes and assign levels
2. Experimental design and construction of choice sets
3. Questionnaire design
4. Analysis of responses

(1) Identify attributes and assign levels

In a DCE, participants are shown a specific number of hypothetical choice scenarios. Each scenario involves the presentation of a decision-making situation, which can be resolved with two or more possible options. The participant is required to make a decision, and select the option which they consider to be most preferable. Options within choice sets are described by levels of specific attributes which are characteristics. They are used within a DCE to evaluate the attractiveness of an option by describing benefits and risks.

When determining the inclusion or exclusion of attributes, it is important to retain the most realistic and plausible attributes which most greatly impact the attractiveness of an option. The selection of attributes can be informed by primary data, e.g. focus groups and interviews, or secondary data, e.g. policy documents and published literature. Generally, three to seven attributes are recommended in a DCE. This is because it has been acknowledged that the greater the number of attributes, the greater the occurrence of compensatory decision-making. That is, as the volume of presented data to the respondent increases, respondents will simplify their decision-making to only select options which present them with the most favourable level of the attribute they consider to be most important.

After deciding which attributes should be included in the DCE, levels (i.e. measurement units) must be assigned to each attribute. They may be quantitative (e.g. time, cost, distance) or qualitative (e.g. ordinal or categorical). The levels should be realistic and plausible. For example, 100% efficacy should not be presented as a level if it is not possible with current treatment options.

(2) Experimental design and construction of choice sets

A DCE systematically varies the attribute levels in order to elicit a behavioural response from which the determinants of choice can be investigated. The respondent will select the most attractive option.

A DCE must adhere to systematic methods of experimental design. This is to cover four main objectives (Health Economics Research Unit, 2010):

i. To estimate the desired forms of utility function (including non-linearity if required)
ii. To ensure the statistical efficiency of the experiment allows for precise estimation of parameters
iii. To not place an excessive cognitive burden on respondents
iv. To ensure realistic choice process and presentation of choices

In order to create options, full factorial designs or fractional factorial designs (FFD) can be used.

(i) Full factorial designs

The total number of possible profiles, i.e. combinations of attributes and profiles for a given number of levels (L) and number of attributes (A) is calculated using the formula L^A. A greater number of attributes and/or levels results in a greater number of possible profiles. If the number of profiles is unmanageably large, it should be reduced to prevent decision-making fatigue which can compromise the validity of responses.

There are three main ways to decrease the number of profiles, (a) reduce the number of attributes or levels, (b) block the design, (c) use a subset of profiles obtained via FFD. Reducing the number of attributes and levels where possible is the first recommendation. If this is not possible or does not result in a sufficiently reduced number of profiles, FFD is the next recommended step. The profiles should not be reduced at random because correlations in the data may prevent model estimation, and multicolinearity may be introduced when there is not enough variation, and variables may move in the same direction which presents difficulties when trying to determine the drivers of preference.

(ii) Fractional factorial designs

Fractional factorial designs (FFD) reduce the total number of profiles from a full factorial design into a subset of all possible combinations of attribute levels.

A good fractional factorial design should result in (Health Economics Research Unit, 2010):

1. Level balance: all levels of each attribute should occur with equal frequency
2. Orthogonality: the levels of each attribute vary independently of each other with minimal correlations. For any two attributes all combinations of pairs of levels appear with proportional frequencies.
3. Minimal overlap: the probability that an attribute level repeats itself in each choice set should be as small as possible. We can achieve this if the difference between the number of times that any two levels of an attribute are replicated is at most one.
4. Utility balance: options within a choice set should be equally attractive to respondents

Designs can be created by statistical software (e.g. SAS), catalogues, websites, and consultations with experts.

Choice sets may either be forced where an alternative must be chosen, or there may be the inclusion of an opt-out/"neither" option which would apply to individuals who refuse to select the alternatives provided.

(3) Questionnaire design

In this stage, the choice sets are represented within a questionnaire. The format of the questionnaire is also decided on, e.g. paper, online, interview with facilitator to guide the participant.

(4) Analysis of responses

The most widely used model for analysing DCE responses is the multinomial logit (MNL). However, alternative analysis methods are also used and include probit, random effects probit, logit, random effects logit, nested logit, mixed logit, and latent class.

The utility derived by an individual (u) is an observable systematic component (v), with an unobservable random component (ε). Essentially, u = v + ε

The information we obtain from a DCE is the observable systematic component:

$$V = \propto + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K$$

α – alternative specific constant (ASC)

X – attributes

β – parameters

From this equation, it is possible to calculate the trade-offs between attributes or marginal rates of substitution which is given by the ratio of attribute coefficients β_x $(x = 1,2, \ldots, K)$.

3.3.3.3 Swing-weighting

MCDA is a decision-making method which can take into account multiple favourable and unfavourable effects simultaneously. It breaks a decision-making scenario into its constituent elements, and derives the value of each element in a piecewise fashion before analysing them to provide a solution.

There are eight steps to performing an MCDA.

(1) Establish the decision context

In this step, the decision question to be evaluated is established.

(2) Identify the alternatives

Potential options, solutions, or actions (referred to as alternatives) which can be taken by the decision maker(s) are identified.

(3) Identify criteria

The favourable and unfavourable effects of treatment (referred to as criteria) are used to construct a value tree, in which a root node branches out to other nodes grouping these effects.

(4) Score the alternatives against the criteria

Once all the criteria have been identified, it is necessary to input into the model how well each alternative performs according to the criteria under consideration, i.e. data on the favourable and unfavourable effects of treatment.

(5) Create value functions

There are three different types of value functions: a) linear: the scores inputted for a criterion are normalised across the 0-100 score proportionately to their values, b) piecewise linear: this is an approximation for normalising continuous data on a non-linear scale at specified linear intervals, c) discrete: these assign values to input scores based on categories of the data via a step function.

(6) Weighting

MCDA compares the value of a change in the amount of one favourable or unfavourable effect criterion with the value of a change in the amount of another favourable or unfavourable effect criterion. At this stage in the modelling, options have already been scored and these scores have been converted to values on the same scale, which allow them to be compared directly to each other. Now the changes or "swings" over these scales are considered by the decision-maker(s) to assign weights.

(7) Analysis

The total weighted score for each alternative is derived from input scores, value functions, and weights for each individual favourable and unfavourable effect criterion, using the value tree. A sensitivity analysis can then be performed to assess how robust the final decision is to variations in the weights assigned to each criterion by the decision-maker(s).

3.3.3.4 MACBETH

MACBETH stands for Measuring Attractiveness by a Categorical-Based Evaluation Technique. It is a decision-making methodology implemented through software, with a key aim to elicit and numerically represent value judgments based on stated criteria (Bana e Costa and Vansnick, 1999). The method measures the value of favourable and unfavourable effect criteria under consideration through non-numerical pairwise comparisons, where decision-makers assign one of seven qualitative categories of difference in value. Using these qualitative judgements, value scores for options and weights for criteria are then derived mathematically.

In total, there are five steps to performing MACBETH. They are: (1) defining criteria, (2) constructing a multi-dimensional scale, (3) inter-criteria evaluation, (4) intra-criteria evaluation, and (5) analysis. Each step is described below.

(1) Defining criteria

The first step is to define a set of criteria necessary for decision-making, i.e. the favourable and unfavourable effect criteria to be incorporated into the MACBETH model.

(2) Construct a multi-dimensional scale

MACBETH constructs a multidimensional performance scale by using the “determinants technique”. This technique involves a three-step process to be achieved after defining criteria and before initiating the weighting process.

This requires two values to be assigned to each criterion:

a) A minimum value, called “neutral”, which represents the value of a criterion at which an alternative would be minimally attractive, but still acceptable. This is not necessarily the minimum value a criterion can take; it is only the threshold value at which it is considered “minimally attractive”.
b) A maximum value, called “good”, which represents the value of a criterion at which an alternative would be satisfactory. This is not necessarily the maximum value a criterion can take; it is only the threshold value at which is it considered satisfactory.

Assigned limits are then used scaled so that the “neutral” value is 0 and the “good” value is 100.

Next, each criterion is labelled as either “determinant” (D), “important” (I), or “secondary” (S). A determinant criterion is pivotal to a decision; therefore, if the performance of an alternative in a determinant criterion is negative, it is a sufficient condition for the alternative as a whole to be considered negative.

Lastly for this step, a reference of good and neutral performance must be defined on the set of criteria. This step requires the DM to determine two reference profiles:

a) A good reference: one where all determinant criteria are satisfactory and a majority of important criteria are satisfactory

b) A neutral reference: one where a majority of determinant and important criteria are neutral, without any criteria being negative

MACBETH uses cardinal value information— where the attractiveness of criteria is not only ordered, but its numerical difference can also be derived. To do this, MACBETH uses a non-numerical pairwise comparison questioning mode which elicits qualitative judgments rather than quantitative ones, from which an interval value scale can be constructed. Pairwise comparisons are made using a qualitative scale that includes seven options: neutral, very weak, weak, moderate, strong, very strong, and extreme.

(3) Inter-criteria evaluation

Pairwise comparisons are made between criteria using the qualitative scale to generate an ordinal, pre-cardinal, and cardinal scale.

a) Ordinal information

Pairwise comparisons of criteria are made. For each pairwise comparison, the decision-maker(s) is/are asked, “Is one of the two criteria more attractive than the other and if yes, which one?”

b) Pre-cardinal information

Using the ordinal information, the criteria are ordered by most to least important. The decision-maker(s) is/are then asked to judge the difference of attractiveness using the qualitative scale of a swing from neutral to good through pairwise comparisons of combinations of criteria. Disagreements or hesitations between two neighbouring categories are allowed. Inconsistencies need to be addressed; judgements need to be consistent.

c) Cardinal information

The qualitative information is not generally sufficient as many possible scales can respect the elicited information. Hence, the decision-maker(s) is/are asked to observe the MACBETH scale axis to compare value intervals. The interval between elements, i.e. the difference in attractiveness, can be adjusted within the limits that respect previous information, to generate a final cardinal scale.

(4) Intra-criteria evaluation

This evaluation involves the comparison of levels of single criteria. Two methods are proposed: direct and indirect evaluation. Generally, only one method is used for each criterion.

a) Direct evaluation

Direct evaluation involves making pairwise comparisons of the performance of alternatives with that of the good and neutral reference within single criteria using the qualitative scale.

b) Indirect evaluation

Indirect evaluation involves building value functions by making pairwise comparisons of different values of criteria using the qualitative scale. It is an indirect technique as the attractiveness of an alternative is calculated based on its performance and the derived value function.

(5) Analysis

The framework provides three means of analysing results: the main, sensitivity, and robustness analyses.

a) Main analysis

The attractiveness of each option is derived based on the cardinal scale, which uses all the information from the weighting process.

b) Sensitivity analysis

It is important to verify how the recommendation for an option would change based on the values assigned to some criteria. The sensitivity analysis can represent the attractiveness of each option as a function of the weight of single criteria. The threshold weight at which the best option changes is an important value to note when considering the uncertainty in the value weights.

c) Robustness analysis

This analysis checks whether the best option changes when taking into account ordinal and/or pre-cardinal intra-criteria and inter-criteria information. This step will inform whether there is “additive dominance” between two alternatives, or if the alternatives show “incomparability”. The first means that one option was globally more attractive than the other, whereas the latter means that neither option is more attractive than the other. Overall scores should measure the relative attractiveness of all the options across all the criteria.

3.3.4 Evaluating the process of involvement

3.3.4.1.1 The perspective of the participant

For each of the methodologies, participants were asked how they felt about the preference elicitation process. It is important to address how we can evaluate the benefit-risk methodologies from a participant’s perspective to encourage future participation and meaningful involvement. Likert questions were developed by the team to ask how participants how they viewed the process of preference elicitation. Respondents rated how strongly they agreed or disagreed with the following statements on a five point scale:

- "It was easy to make comparisons between the outcomes."
- "The questions adequately reflect the aspects of relapsing remitting MS that I feel are important."
- "Enough information was provided, in a clear and understandable format, to enable me to answer the questions."
- "I would be happy to take part in similar surveys in the future."

In addition to this, there were two questions where respondents had the option to provide free text responses:

- Were any of the questions particularly difficult to answer? Please give details.

- Please provide any additional comments or suggestions in the box below.

Lastly, participants were also invited to provide any additional free text comments or suggestions to improve the preference elicitation methodology.

3.3.4.1.2 The perspective of the facilitator

Although frameworks to address PPI already exist, they predominantly focus on PPI in general, or on the context of health research and health services. There is not a clear framework to guide the PPI in the benefit-risk assessment of medicines. To address this, the principles and indicators of PPI were extracted from key documents and carefully reviewed by the team to create a framework to guide the application, reporting and evaluation of PPI in the benefit-risk assessment of medicines (Table 3-2).

Table 3-2 framework developed to guide the application of PPI to the benefit-risk assessment of medicines and regulatory decision-making

	Step	Points to consider
Steps to be addressed during planning	**Determine the purpose of PPI**	What is the aim?
		Which benefit-risk assessment methodology(/ies) are going to be used?
		During which stage(s) of the methodology(/ies) is involvement required and/or desired?
		What is the desired level of involvement for each stage?
	Ethical approval	Is ethical approval required? Who needs to approve ethics application? How long is needed to obtain ethics approval?
	Conflict of interest declaration	Do the researcher(s) have any potential conflicts of interest? Do the patient organisation(s) and/or the patients have any potential conflicts of interest?
	Address potential barriers and negative outcomes from (a) a patient and public involvement perspective, and (b) a benefit-risk methodology perspective	What are the potential barriers to meaningful involvement?
		What are the potential negative outcomes of involvement?
		How can the barriers and negative outcomes be alleviated?
	Training	Do participants require training and support? If so, how will this be addressed?
		Do researchers require training and support? If so, how will this be addressed?
	Recruitment	Which group of participants will be used to represent patients and the public?
		What is the sample size required for the methodology? How many participants will be recruited?
		What are the methods and anticipated time scales for recruitment?

	Step	Points to consider
	Design a participant information sheet	Is it possible to provide full disclosure of the benefit-risk methodology being studied and the role and anticipated value of patient involvement?
		What are the roles and responsibilities of both the researcher and the participant?
		Do the participants know all of the confidentiality, anonymity, drop-out, and acknowledgement policies?
		What are the anticipated time scales for involvement activities?
	Patient involvement activities	What is the method of communication?
		What is the location of involvement activities (if applicable)? Are there any special considerations, e.g. wheelchair accessibility?
		Are there finances in place to specifically support involvement?
	Finances	Can participants receive adequate financial support for their expenses and contribution?
		What were found to be the positive outcomes of involvement?
Steps to be addressed during evaluation	**Reporting of outcomes**	What were the negative outcomes of patient involvement?
		Did conflicting perspectives or disagreements occur?
	Reporting of conflicting perspectives	At which stage of the process did they occur?
		Who did they occur between?
		What were the different perspectives?
		How were they resolved?
		Is it necessary for participants to be periodically informed of the decision-making process as it progresses?
	Dissemination	How will participants be informed about the results of their involvement?
		How will participants evaluate the processes that they were involved in?
		How will the contribution of participants be explicitly acknowledged?
		How will the involvement process be reported to all stakeholders?
		How will the overall impact of the decision on patients be evaluated?

The decision context of natalizumab: Planning and data preparation

3.3.5 Specification of the decision problem

(a) The decision problem

The aim of the decision problem is to decide whether the emerging risk of PML in the post-marketing period shifted the benefit-risk balance of natalizumab from positive to negative.

Natalizumab received marketing authorisation from the United States Food and Drug Administration (FDA) from November 2004 to February 2005 for the indication of relapsing remitting multiple sclerosis (RRMS). The marketing authorisation was subsequently suspended due to the occurrence of PML but was later reintroduced to the market with strict risk minimisation measures in June 2006—demand from patient organisations lead to its reintroduction. In the EU, the EMA granted marketing authorisation for natalizumab in June 2006. In 2009, the benefit-risk balance of the treatment was reassessed by the CHMP due to new reported cases of PML; the marketing authorisation was maintained with risk minimisation measures in place.

(b) Perspectives

In this assessment, the decision-makers are patients with multiple sclerosis deciding the importance of outcome measures for themselves.

Although many comprehensive, structured and systematic methods to evaluate and trade-off benefits and risks exist within a multitude of decision-making settings, there is a lack of research which formally tests, evaluates and compares formal methods of eliciting patient preferences within a regulatory context. This work plans to address this by testing the feasibility of swing-weighting, MACBETH, AHP and DCE to elicit preferences from patients and the public affected by RRMS.

(c) Comparators

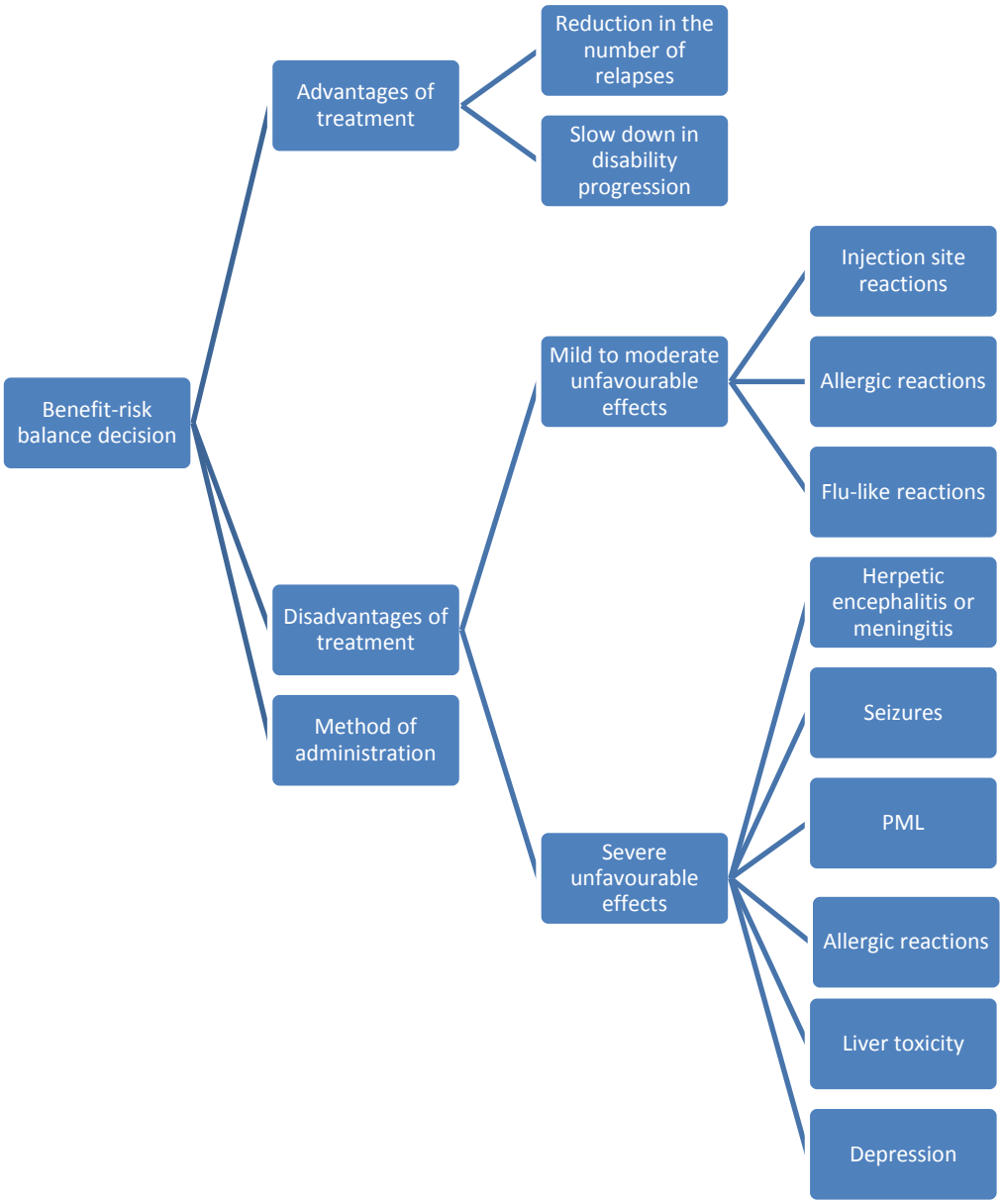
The benefit-risk balance of natalizumab (dose: 300mcg) will be compared against interferon beta-1a (dose: 30mcg), glatiramer acetate (dose: 20mg) and placebo.

This assessment intends to evaluate the benefit-risk balance of natalizumab from a patient perspective. It aims to replicate real life; patients are likely to consider a range of active treatment options which include alternative medicines for the same indication.

(d) Favourable and unfavourable outcomes

The favourable outcome measures to be included in the assessment are: (1) reduction in the number relapses, (2) slowdown in disability progression. The risk measures to be included in the assessment are: (1) injection site reactions, (2) mild to moderate allergic reactions, (3) flu-like reactions, (4) herpetic encephalitis or herpetic meningitis, (5) seizures, (6) PML, (7) serious allergic reactions, (8) liver toxicity, (9) depression.

The value tree used to represent these outcomes is displayed in Figure 2.


Benefit-risk balance decision
Advantages of treatment
Reduction in the number of relapses
Slow down in disability progression
Disadvantages of treatment
Mild to moderate unfavourable effects
Injection site reactions
Allergic reactions
Flu-like reactions
Severe unfavourable effects
Herpetic encephalitis or meningitis
Seizures
PML
Allergic reactions
Liver toxicity
Depression
Method of administration

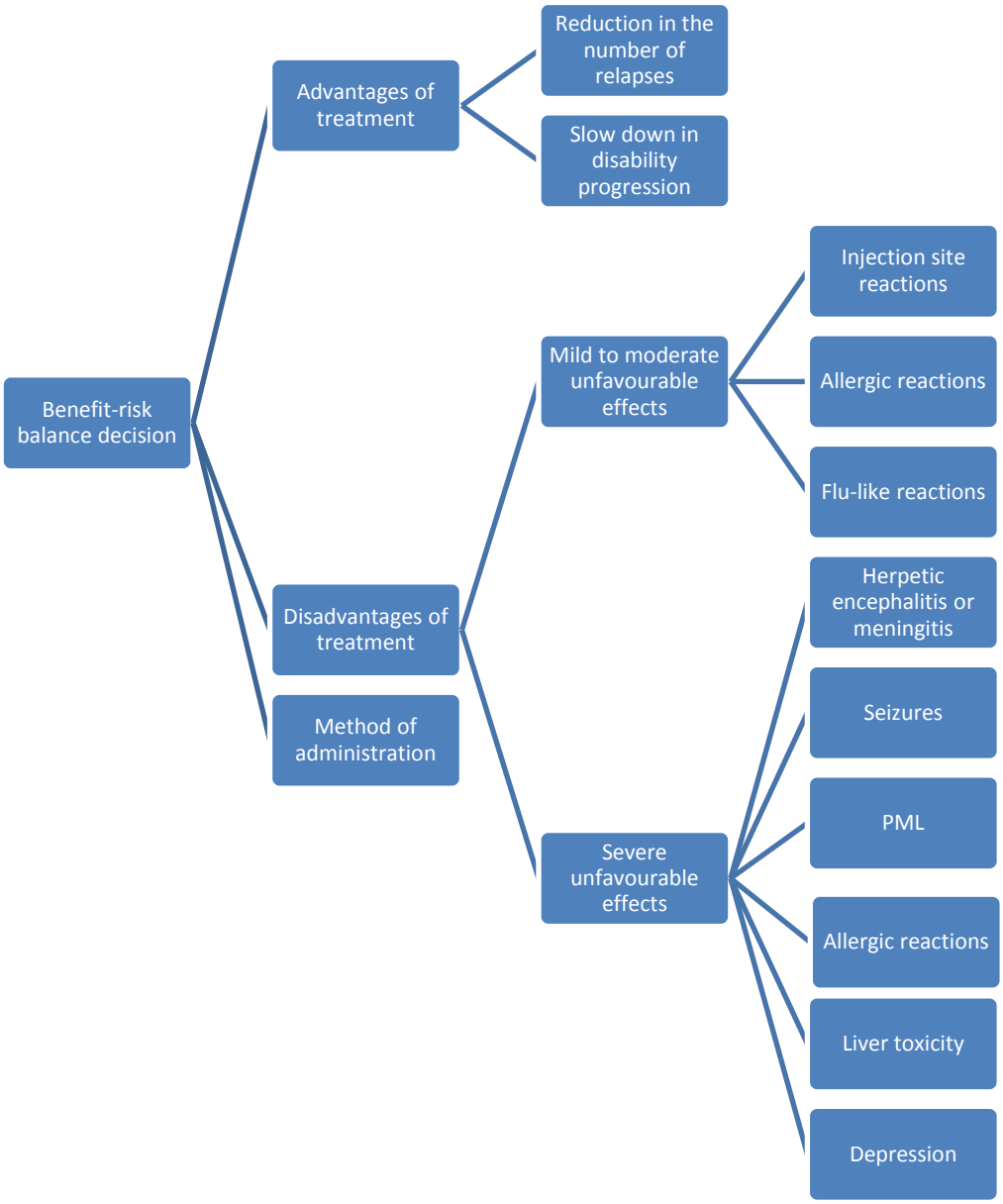


Figure 2 Value tree adopted by the PPI team

(e) Sources of evidence

Evidence regarding phase three clinical trials and post-marketing surveillance will be extracted from a number of key documents: European Public Assessment Reports, Periodic Safety Update Reports, Important Safety Information, Prescribing Information, Summary of Product Characteristics and key academic journal publications.

These sources of evidence were selected as they are believed to provide: (a) a comprehensive picture of the benefits and risks of each treatment, (b) data of a reliable quality, (c) representative benefit and risk information communicated to a range of stakeholders (e.g. physicians are the target audience for Prescribing Information documents, patients are the target audience for the medication package insert included in the Summary of Product Characteristics).

(f) Time horizon

The time horizon for measuring the occurrence of the benefits and risks will reflect those reported in the sources of evidence.

(g) Resources

Within the PPI team there is clinical, regulatory, patient organisation, industry, statistical and academic expertise. The PPI work will be completed within the timeframe of IMI PROTECT WP5.

3.3.6 Development of the value tree

As the PPI work progressed, we had many discussions regarding the value tree. These discussions covered the:

- Inclusion and exclusion of outcome measures
- Imprecise use of terminology
- Unclear representation of outcome severities
- Unbalanced representation of the three active treatments

Although the original value tree was adopted by the Wave One and Wave Two natalizumab case study teams, its applicability within the context of patient and public involvement and preference elicitation was queried. As a consequence, we reviewed options to: (a) amend the current terminology, (b) modify and/or remove current outcomes, and/or (c) include new outcomes. However, it should be noted that there is no such thing as a "correct" or "incorrect" value tree for any benefit-risk decision-making scenario; how the value tree is defined depends on the specification of the decision problem and context. The decision problem and context specified by the PPI team was different to the earlier natalizumab case study teams and so the PPI team adopted a different approach to defining the value tree.

There are an extremely large number of outcome measures reported in publicly available regulatory documents and peer-reviewed publications. It is likely that the outcome measures represented in the benefit-risk assessment may vary according to whose perspective(s) are included. The relevancy of outcomes in the Wave One and Wave Two case study were debated. It was considered desirable for the outcomes to be: (a) more medically relevant, (b) more understandable for patients, (c) more likely to have an impact on benefit-risk assessment from the perspective of a patient. With this in mind, a value tree modified from the Wave One and Wave Two natalizumab case studies was proposed. The original value tree from Wave Two is presented in Figure 3. The measures of unfavourable outcomes versus new proposed measures of unfavourable outcomes are presented in Section 3.3.6.1.

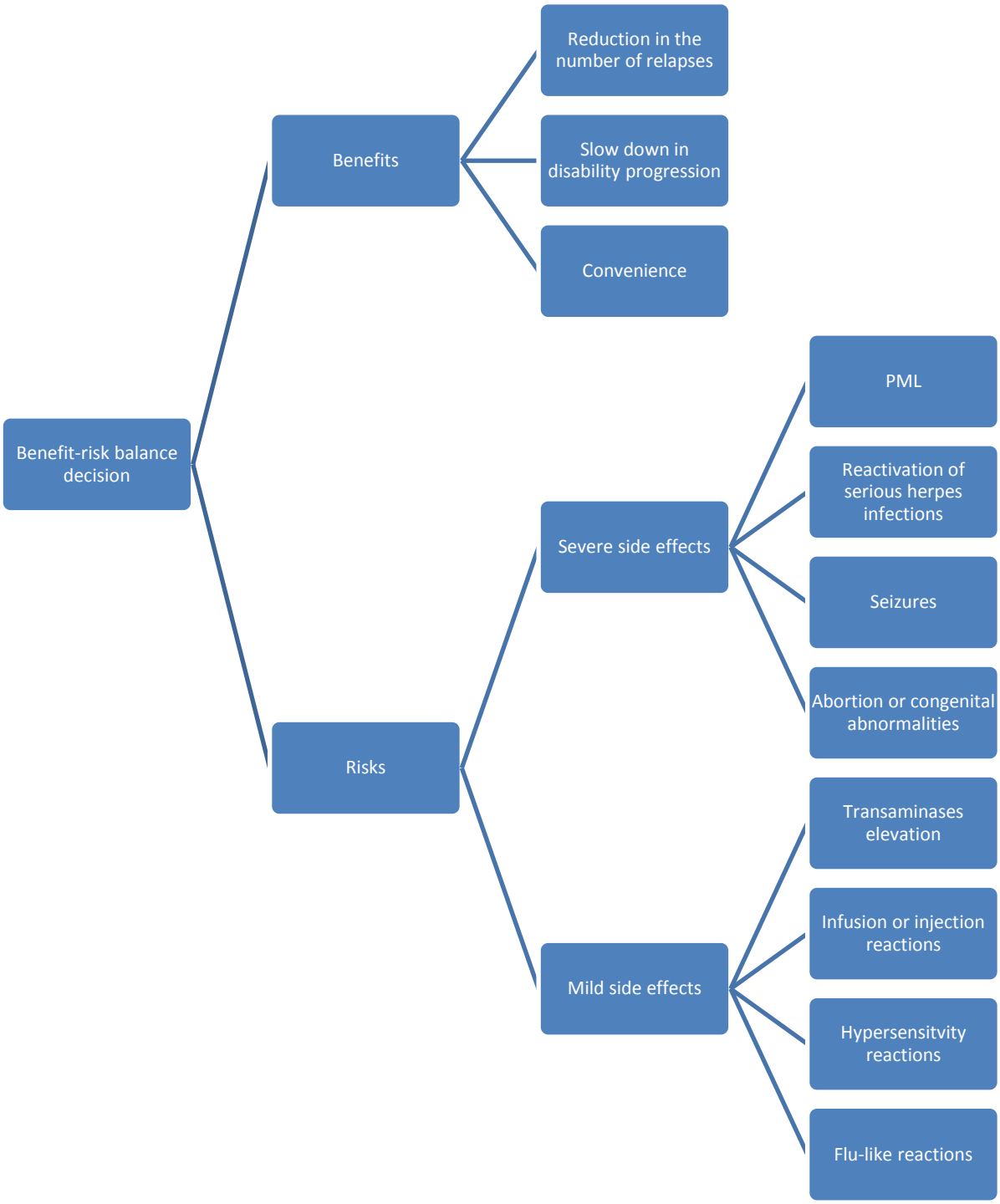


Figure 3 Value tree presented in the wave two natalizumab case study for the purposes of preference elicitation

3.3.6.1 Summary of changes to the value tree

The two categories were renamed: “less serious unfavourable effects” was replaced with common “mild to moderate unfavourable effects”; “more serious unfavourable effects” was replaced with “severe unfavourable effects”. The rationale when naming the original categories was to place them on two separate ends of a “seriousness” scale. However, within regulatory benefit-risk assessment careful use is required when using the word “serious”; there were concerns over an imprecise use of terminology. Although a lay definition of serious was intended to describe the categories, serious has a very specific widely accepted definition within a clinical trial and regulatory context and the names of the categories were changed to reflect this.

Ideally, the patient perspective should have been sought to assign each risk to either the “Mild to moderate unfavourable effects” or “severe unfavourable effects” adverse event categories, or alternatively sought to provide buy-in for the resultant value tree. However, this was not possible due to time constraints. Instead, the categorisation of risk was achieved via a consensus of the WP5 PPI team, and included the perspective of a physician with clinical experience of the conditions.

The outcomes remained the same in the “mild to moderate unfavourable effects” category except “transaminases elevation” was renamed to “liver toxicity” and moved to the “severe unfavourable effects” category. For patients to provide valid weights, it is important that they clearly understand the meanings of the outcomes; it was believed that modifying this term would enhance participant understanding as instead of abstractly referring to biomarkers, the term now refers to real life symptoms more comprehensible to patients. Hypersensitivity reactions were also renamed to “allergic reactions”. This is because when reading treatment information targeting patients (e.g. patient package inserts, summary of product characteristics) “hypersensitivity reactions” are frequently referred to as “allergic reactions”; once again the term is more easily understood by lay audiences.

In the “serious unfavourable effects” category, the outcome miscarriage and congenital abnormalities was removed. This is because there was insufficient data—and therefore insufficient evidence to demonstrate the impact of the three active treatment comparators on pregnancy (although it is important to note that the lack of evidence results in a warning for pregnant women not to use the treatments). The outcome “reactivation of serious herpes viral infections” was redefined to only refer to herpetic meningitis/herpetic encephalitis– one of the important safety concerns related to natalizumab. Once again, it was hoped that limiting the conditions covered by the outcome would enhance participant comprehension and therefore help improve the validity of elicited weights. Two additional outcomes were added with a similar thought in mind: serious hypersensitivity and serious liver problems or damage. Originally, these two outcomes were included under “less serious unfavourable effects” although they were defined with very broad ranges of severity—all the way from mild and transient to severe and permanent, which may have resulted in confusion when asking participants for their preferences. In the new proposed value tree, these two outcomes have two representations of severity included under different nodes. The last outcome added was depression and/or suicide ideation. This change was to ensure adequate representation of interferon beta-1a; it is one of the main safety concerns related to the treatment and is considered to be very important from the patient perspective in a clinical setting when deciding between treatments.

3.3.6.2 Challenges we faced in the development of the value tree

PROTECT benefit-risk case studies primarily focussed on the testing of the methodologies rather than the accuracy of selecting: (a) which outcomes to account for, and (b) which measures should be used to provide the necessary data to represent the outcomes; however, for good practice we noted a number of specific issues which should be addressed when conducting benefit-risk assessment. The points presented here describe several issues we encountered during the development of the value tree:

a) Selection of outcomes

There were disagreements within the team regarding which outcomes we should represented in the value tree. This may be due to how different stakeholders might find different outcome measures important: for example, clinicians understand scientific terminology and biological mechanisms whereas patient advocates can better relate to how symptoms can affect everyday life. We resolved this situation by ensuring that the wide range of knowledge, opinions and perspectives in the PPI team were shared and the process was iterative until a common understanding and consensus was reached.

b) Selection of measures

Outcomes are selected because they are viewed as extremely important. However, there may be many different ways the outcome could be represented through clinical measures. For example, the PPI team thought that it was important to account for relapses. However, when looking at clinical trial data, many different measures could be used to represent this outcome: the proportion of relapse free patients or an annualized relapse rate, or percent decrease in relapse rate compared to placebo. Another example is the outcome is "hypersensitivity reactions"; it could be represented by clinical measures such as hypersensitivity, allergic reaction, anaphylactic or anaphylactoid reaction, urticarial, allergic dermatitis, and/or hives. We adopted the approach where we listed and critically evaluated all available measures which represented the outcome of interest and then selected which measure we believed best reflected the perspective of the decision-maker, with advice from clinical, statistical, and the patient organisation IAPO to ensure that the measures were justifiable and traceable.

a) The number of outcome measures which can be included

It was difficult to determine how many outcomes should be included in the preference elicitation exercise. Including a large number of outcomes results in a more comprehensive representation of the decision-making scenario. However, there may be limitations on how many can be outcomes can be included; this may vary by which methodologies are adopted and/or the feasibility of asking an increased number of preference elicitation questions. To resolve how many outcomes should be represented for natalizumab we were specific when defining the decision problem and only included those which reflected the situation from the required perspective after a balanced evaluation.

b) The degree of severity

Each outcome and/or measure may represent a broad range of severities: from mild to severe and transient to permanent. One example is the outcome of "injection site reaction", as this can range from temporary reddening and bruising to permanent tissue necrosis which requires hospitalisation. We found two potential solutions to addressing how severity is accounted for: (a) represent the severity which best matches up to the adverse event resulting from the treatment (although there may be difficulties when using comparators), and (b) represent the whole range of severities which is implied by the term. Ultimately however, for our decision scenario it was best to limit the range of severities represented by the measure as much as possible and clearly define the level of severity when eliciting preferences. This was to avoid confusion, simplify trade-offs and ensure the scientific validity of the findings.

c) Availability and quality of data

We used publicly available outcome measure data published in regulatory documents and academic journal articles. Unfortunately, we experienced a relatively large amount of missing, imprecise, unclear, and even in some cases inconsistent data. Ideally, the outcome measure data used in benefit-risk assessment should be precisely defined,

reliable and accurate; although this was found to be relatively achievable with clinical trial data, data collected during post-marketing surveillance represented a challenge.

Table 3-3 Challenges with publicly available clinical trial and post-marketing surveillance data

Challenges with publicly available clinical trial data	Challenges with observational post-marketing surveillance data
There may be: • Unclear inclusion/exclusion criteria • Varying timeframes/periods of exposure when comparing across trials • Only a subset of the outcome measures that were collected may be reported • Summary statistics are typically reported rather than full patient level dataUnclear effective allocation concealment and blinding • Unclear whether randomisation to treatment groups was effective • Trials are powered for efficacy and not safety	There may be: • Unclear baseline population characteristics • Missing denominators when calculating rates • Lack of clarity to understand if the outcomes are associated with/caused by the treatment • Unclear, unknown, and/or varying timeframes of exposure • Differences in information/data collected by databases used in observational studies • Difficulty in assessing the impact various biases and measured/unmeasured confounding may have on the study results • Unclear generalisability of results of studies conducted in specific populations • Difficulty in establishing whether there is a causal relationship between the exposure and the adverse event

Some of these issues may be resolved in a regulatory setting as non-publicly available, non-published data may have been available to those making the decision; it is likely that the clinical trial and post-marketing surveillance data which regulators view is much more comprehensive than the data we used for natalizumab. If systematic methods of benefit-risk assessment are applied to these data in this setting, they are likely to produce results with increased validity. However, if industry is to apply formal methods of benefit-risk assessment, there may still be issues; it might be possible to use comprehensive data for their product, but they will not be able to use similar quality outcome data if they wish to use a comparator, i.e. compare their treatment against alternative medicines manufactured by different companies used to treat the same indication. It is important to develop statistical methods of handling data of different quality, timeframes, and sources for use in benefit-risk assessment. Another IMI project called GETREAL is looking into this topic.

d) A clear audit trail

We found that performing a benefit-risk assessment is a large project, and it was necessary to break it down into a number of smaller tasks whereby smaller subteams or individuals completed specific pieces of work. Although it was clear to those who completed the smaller tasks how necessary but implicit decisions had been made, there was confusion from the wider group when trying to trace back certain pieces of information or knowledge. This was partially due to the vast numbers and complexities of the interdependent tasks, in addition to how team members represented a range of different backgrounds, i.e. what is clear to people from one background (e.g. clinician) might

not be clear to one from a different background (e.g. statistician). For example, lots of different measures may be available to represent an outcome: the choice of which measure is selected may be critical. One individual might find that it is extremely clear which measure to select, whereas another individual might not agree and/or understand how the decision was made. Our proposed solution was to create and maintain a clear reporting framework that specified which interim decisions have been made throughout the benefit-risk assessment process, and note the criteria and any background information used to guide the decision.

We experienced many disagreements regarding which outcomes should be included in the natalizumab benefit-risk assessment model after extracting and reviewing all of the publicly available outcomes from regulatory documents and key study publications. Following discussions, it was found that much of this was attributable to the varying professional perspectives of a multidisciplinary team consisting of industry, clinicians, statisticians, academic, and patient organisation partners. Our proposed solution to handle this was to ensure that knowledge, opinions and views were shared and the process made iterative until consensus was reached. Also, we believe that there should be clear justification regarding whether an outcome was selected or not and why specific measures were selected to represent the outcome.

e) Difficulties accounting for three active comparators

Clinical trials traditionally compare a single active treatment versus a placebo to determine efficacy and safety. However, the application in this study varies: we wished to include multiple active comparators. Multiple active comparators were considered to better represent the real-life decision-making scenarios faced by patients with their clinicians, i.e. they do not select between a single treatment and no treatment; they select either no treatment or an active treatment from a range of treatments which have received marketing authorisation for their specific indication. Natalizumab, interferon beta-1a, and glatiramer acetate are three comparators which we included to represent three realistic treatment options for patients with RRMS and all have a distinctly different benefit-risk profile.

As previously discussed, although theoretically any number of outcomes can be included in many types of systematic benefit-risk assessment methodologies, there are feasibility issues. A larger number of outcomes included in the methodology will result in a larger number of preference elicitation questions which can be burdensome on participants and require a great amount of time and cognitive demand. The number of outcomes which we included in the methodologies became even more difficult to manage as we wished to account for multiple active comparators. All three active treatments under consideration have different favourable and unfavourable effect profiles.

Naturally as the number of comparators increase, the number of outcomes too will increase. It is not possible to include the most prevalent or important outcomes from every single treatment or else the total number of outcomes for all treatments may be extremely large once these are all compiled into a single value tree. It is important to ask which should be included or excluded based on the decision context and adopted perspective. To address this, firstly we (a) compared the different profiles, and (b) compared the outcomes listed versus the ones included in the proposed update to value tree through examining our pre-specified data sources. Next, we included outcomes based on perceived relevance, i.e. what we considered to be relevant from a patient, regulatory and medical perspective. Then, the remaining outcomes were moderated from a decision-analyst perspective, i.e. (a) how many outcomes can feasibly be included, and (b) pre-emptively consider what is likely to tip the balance and influence the determination of the benefit-risk balance.

We wished to build a fair and balanced picture which would represent the important issues associated with each of the three treatments under consideration: we did not wish to over represent the favourable or unfavourable effects

from a specific treatment (i.e. include a disproportionate number of AEs or efficacy measures from one treatment) which could potentially bias the assessment. To accomplish this we aimed for an adequate and balanced representation of the three treatments from the team perspective.

3.3.7 Data

3.3.7.1 Data extraction

Data were extracted from: European Public Assessment Reports (EPARs), Important Safety Information, Summary of Product Characteristics, Prescribing Information, and key academic publications.

a) European public assessment reports

A number of regulatory documents produced by the EMA are publicly available online (http://www.ema.europa.eu/ema/), or by request for those interested in regulatory processes and authorisation decisions. EPARs often contain clinical trial and regulatory post-marketing surveillance data. The "product information" and "scientific discussion" documents are also available to view online on the EMA website. Only data for natalizumab were extracted from EPARs; this was due to the extremely large amount of time required to go through and extract information from the vast amounts of information presented in heterogeneous formats.

Over 110 different outcome measures were presented.

b) Important Safety Information

The Important Safety Information is a document authored by industry and/or sponsor and/or marketing authorisation holder targeted at physicians, patients and the public. The efficacy and safety outcomes included in the document are based on the strength of evidence representing the outcome and a duty to inform.

Table 3-4 displays the outcomes described in the "Important Safety Information" for each of the active treatments considered by the PPI team.

Table 3-4 Outcomes described in the Important Safety Information for natalizumab, glatiramer acetate and interferon beta-1a

	Natalizumab	Glatiramer acetate	Interferon beta-1a
Favourable effects	Reduce flare-ups. Slow down the progression of physical disability. Reduce new brain lesions and lesion activity.	Reduce relapses. Reduce relapse rate at two years. Reduce the development of some types of brain lesions.	Decrease the number of flare-ups. Slow the occurrence of some of the physical disability that is common in people with MS.
Unfavourable effects	PML* Liver damage* Herpes encephalitis and meningitis* Allergic reactions* Unusual or serious	Immediate post-injection reaction* Chest pain Lipoatrophy and necrosis Potential effects on immune response	Behavioural health problems including depression, suicidal thoughts or hallucinations* Liver problems, or worsening of liver problems including liver failure and death*

	infections Serious allergic reactions* Common: Headache Nausea Feeling tired Urinary tract infection Joint pain Lung infection Depression* Pain in your arms and legs Diarrhea Vaginitis Rash Stomach area pain Other: Nose and throat infections		Serious allergic reactions and skin reactions* Heart problems, including heart failure Blood problems Seizures* Infections Thyroid problems. Common: Flu-like symptoms*

*Outcome included in PPI value tree

When looking at the favourable effects represented in the Important Safety Information, all three treatments are reported to reduce relapses, natalizumab and beta-interferon are reported slow physical disability, and natalizumab and glatiramer acetate are reported reduce brain lesions. The favourable effects included in the Wave One and Wave Two natalizumab case study were to (a) reduce relapses and (b) slow physical disability: the PPI team decided to retain these as outcomes. A reduction in brain lesions was not added to the favourable effects because its regulatory/medical relevance questioned. Inclusion of this outcome may have resulted in double counting because it is correlated with other outcomes which would subsequently result in an overestimation of benefit. The outcome would have also been difficult to explain, i.e. it would be challenging to describe how the physiological effect of brain lesions directly translates onto the patient experience of the treatment and making benefit-risk decisions which in turn may result in inaccurate or invalid weights.

When looking at the unfavourable effects represented in the Important Safety Information, it is clear that hypersensitivity reactions are relevant to all three treatments; they were reported under the headings of allergic reactions, immediate post-injection reaction, and serious allergic reactions. Therefore, both hypersensitivity and serious hypersensitivity reactions were represented in the PPI value tree.

For natalizumab, the outcome of PML was retained in the value tree because it is the outcome which prompted the benefit-risk assessment scenario represented in the Wave One and Wave Two case studies according to the EPARs.

For glatiramer acetate, the outcomes of lipotrophy and necrosis were considered—the outcomes are severe forms of injection-site reactions. However, they do not have a high probability of occurrence and because of this the team represented this outcome in the value tree via the broader category of injection-site reactions.

For interferon-beta 1a, the outcome of depression was included and liver problems were represented (both are also relevant to natalizumab). Seizures and flu-like symptoms were also retained from the original value tree.

c) Summary of Prescribing Characteristics and Prescribing Information

An important source of favourable and unfavourable effects information to physicians and RRMS patients when making decisions about treatments is the Summary of Product Characteristics (SPC). The frequencies of unfavourable effects are expressed in patient-years using the definitions of very common (>1/10), common (≥ 1/100 to < 1/10), uncommon (≥ 1/1,000 to < 1/100), rare (>1/10000 to <1/1000), very rare (<1/10000) and not known (cannot be estimated from the available data). Data on the unfavourable effects of treatment considered by the team were extracted from the SPCs and are presented in Table 3-5. For natalizumab, the frequencies of adverse events were not reported in the SPC if they did not have an, "incidence of 0.5% greater than reported with placebo". For glatiramer acetate, data on the unfavourable effects of treatment were reported if they were more frequently reported in glatiramer acetate vs. placebo-treated patients from three pivotal clinical trials. The data for interferon beta-1a was identified, "from studies (clinical trials and observational studies, with a period of follow-up ranging from two years to six years) and other adverse reactions identified through spontaneous reporting from the market, with unknown frequency".

Table 3-5 Outcomes described in the Summary of Product Characteristics for natalizumab, glatiramer acetate and interferon beta-1a

Outcome in the value tree	Term used in SPC	Natalizumab	Glatiramer acetate	Interferon beta-1a
Injection/infusion site reaction	Injection site reaction	-	very common	not known
Mild common allergic reactions	Hypersensitivity reactions	uncommon	common	not known
Flu-like symptoms	Flu-like symptoms	-	-	very common
Herpetic encephalitis and herpetic meningitis	Unusual (opportunistic) infections	Rare	-	-
PML	PML	uncommon	-	-
Seizures	Seizures Convulsions	-	 rare to very rare	not known
Serious allergic reactions	Anaphylactic reactions	-	rare to very rare	not known
Serious hepatic events	Hepatomegaly Hepatic failure	-	uncommon	 not known
Depression/suicide ideation	Depression Suicide attempt	-	very common uncommon	common not known

When cross-referencing the outcomes described in the SPCs with the PPI value tree, the relevancy of each outcome in relation to each treatment becomes clear (Table 3-6).

Table 3-6 Outcomes in the value tree and relevancy to each treatment

Category	Outcomes	Relevancy to treatment(s)
Common mild to moderate unfavourable effects	Infusion or injection site reactions	All three

	Allergic reactions	All three
	Flu-like reactions	Interferon beta-1a
	Abnormal liver function	Interferon beta-1a
Serious unfavourable effects	Herpetic encephalitis	Natalizumab
	Seizures	Interferon beta-1a
	PML	Natalizumab
	Serious hypersensitivity	All three
	Serious hepatic events	All three
	Depression and/or suicide ideation	Interferon beta-1a

d) Key publications

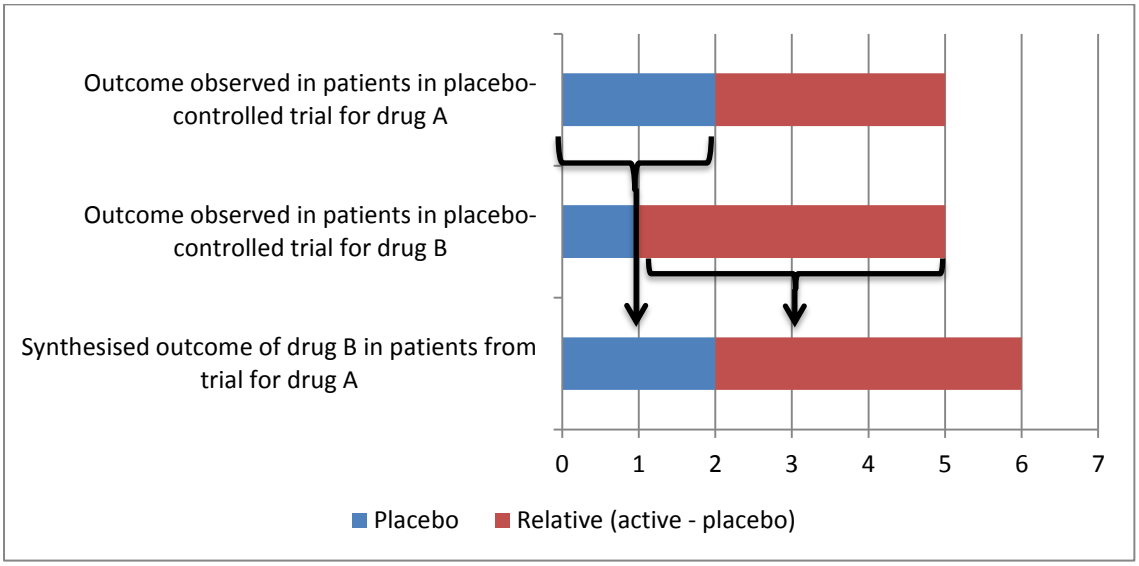
Data were extracted from a number of key academic publications identified by the case study team:

- Polman, C., O'connor, P., Havrdova, E., Hutchinson, M., Kappos L, Miller, D.H., Phillips, J., Lublin, F., Giovannoni, G., Wajgt, A., Toal, M., Lynn F, Panzara, M.A. and Sandrock, A.W. (2006). A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine,* **354**, 899-910.
- Jacobs, L., Cookfair, D., Rudick, R., Herndon, R., Richert, J., Salazar, A., Fischer, J., Goodkin, D., Granger, C., Simon, J., Alam, J., Bartoszak, D., Bourdette, D., Braiman, J., Brownscheidle, C., Coats, M., Cohan, S., Dougherty, D., Kinkel, R., Mass, M., Munschauer, F., Priore, R., Pullicino, P., Scherokman, B. and Whitham, R., and the Multiple Sclerosis Collaborative Research Group (1996). Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Annals of Neurology*, **39**, 285-294.
- Johnson, K.P., Brooks, B.R., Cohen, J.A., Ford, C.C., Goldstein, J., Lisak, R.P., Myers, L.W., Panitch, H.S., Rose, J.W. Schiffer, R.B. and the Copolymer 1 Multiple Sclerosis Study Group (1995). Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial.. *Neurology,* **45**, 1268-1276.
- Kappos, L., Radue, E., O'connor, P., Polman, C., Hohlfeld, R., Calabresi, P., Selmaj, K., Agoropoulou, C., Leyk, M., Zhang-Auberson, L. and Burtin, P. (2010). A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *New England Journal of Medicine*, **362** 387-401.
- Bloomgren, G., Richman, S., Hotermans, C., Subramanyam, M., Goelz, S., Natarajan, A., Lee, S., Plavina, T., Scanlon, J.V., Sandrock, A. and Bozic, C. (2012). Risk of natalizumab-associated progressive multifocal leukoencephalopathy. *New England Journal of Medicine*, **366**,1870-80.
- FDA prescribing information for beta-interferon
- FDA prescribing information for Glatiramer acetate
- Summary of Product Characteristics for Glatiramer acetate

3.3.7.2 Data synthesis

For preference elicitation purposes, clinical outcomes must be expressed in absolute, not comparative, measures. For example, it is possible to attach a subjective value to an X% risk of experiencing a given adverse effect. But it is not possible to place a value on a relative risk of the same event of Y% (compared to some baseline), without also being given the baseline rate. For this reason, an absolute measure of incidence of each benefit/risk is required for

all the treatments under consideration. However, a direct comparison of the absolute outcomes observed in different studies would be susceptible to bias due to heterogeneities in the underlying rates seen in the different trial populations. Therefore relative measures of each outcome are extracted from each trial, expressing the difference in the absolute rate between the active and placebo arms. These relative measures are assumed to be robust to changes in populations. By combining the relative measures with a single baseline rate appropriate to the target population (i.e. those patients to whom the benefit-risk assessment applies), an outcome measure on the absolute scale can be synthesised for each treatment and these can be compared fairly. This is illustrated in the chart below.



In this case, the placebo group from the natalizumab was selected as the baseline population and the risk difference was the relative measure adopted for most outcomes. However, exceptions may be required because of heterogeneities in the data, and pragmatic adaptations may be applied to the model on a case by case basis.

Estimates of the variance of the outcome measures will be obtained by Monte Carlo simulation using a Bayesian implementation of the data synthesis model. The key data relevant to the outcomes represented are presented in Table 3-7.

Table 3-7 Table displaying final data used by the PPI team

GROUP	CATEGORY	OUTCOME	MEASURE	DRUG	MEAN OUTCOME
Benefit	Relapse	Relapse	2 year relapse rate	Placebo	1.46
				Natalizumab	0.46
				Interferon	1.19
				Glatiramer	1.04
	Disability progression	Disability progression	6-month confirmed % progressing after 2 years	Placebo	0.23
				Natalizumab	0.11
				Interferon	0.14
				Glatiramer	0.18
	Convenience	Convenience	Route and frequency of adminsitration	Placebo	oral od
				Natalizumab	iv qm hosp
				Interferon	im qw
				Glatiramer	sc od
Risk	Infection	Herpetic encephalitis		Placebo	0.0%
				Natalizumab	0.002%
				Interferon	0.0%
				Glatiramer	0.0%
		PML		Placebo	0.0%
				Natalizumab	0.213%
				Interferon	0.0%
				Glatiramer	0.0%
	Liver Toxicity	Transaminases elevation	ALT >5x ULN	Placebo	4.0%
				Natalizumab	5.0%
				Interferon	4.0%
				Glatiramer	4.0%
	Neurological	Seizures		Placebo	0.0%
				Natalizumab	0.0%
				Interferon	3.0%
				Glatiramer	0.0%
	Others	Injection site reactions		Placebo	0.0%
				Natalizumab	0.0%

				Interferon	12.5%
				Glatiramer	90.0%
		Allergic/Hypersensitivity Reactions			
				Placebo	0.0%
				Natalizumab	4.0%
				Interferon	0.0%
				Glatiramer	1.0%
		Flu-like reactions		Placebo	39.9%
				Natalizumab	39.9%
				Interferon	60.8%
				Glatiramer	39.9%
		Serious allergic reactions		Placebo	0.0%
				Natalizumab	0.5%
				Interferon	0.0%
				Glatiramer	0.0%
		Depression		Placebo	12.5%
				Natalizumab	14.8%
				Interferon	12.5%
				Glatiramer	12.5%

3.4 Research procedure

3.4.1 Sample selection

In the definition of "patient and public" which we adopted, an extremely large range of individuals are represented. When deciding which individuals we should include for participation our study, we discussed each of the groups in turn—reflecting on our personal thoughts and feedback we had received from other professionals at conferences and meetings. Our discussions are displayed in Table 3-8, and we acknowledge that many of the advantages and disadvantages are common to all stakeholders.

Table 3-8 Perceived advantages and disadvantages of selecting different groups for preference elicitation activities

Stakeholder group	Advantages	Disadvantages
Clinical trial participants	• If industry is engaging in a clinical trial these groups of individuals are easy to access • The preference elicitation activities can easily be incorporated into the overall trial activities	• These individuals may not be truly "representative" of the patient perspective; the nature of the fact that they would be willing to participate in a clinical trial means that they may be willing to take on a greater level of risk than the "average" patient.
Patients	• These individuals make real life treatment decisions for themselves. • They understand the burden of disease and its impact on quality of life from a first-hand perspective—in addition to current treatments.	• These individuals may find it difficult to participate in preference elicitation activities because of the burden of their disease, e.g. physical disabilities may make them unwilling/prevent them from travelling to a decision conference.
Potential patients	• Differences in disease severity (newly diagnosed vs. chronically ill patients) allows for a wide range of preferences to be potentially collected.	
Disabled people		• May have difficult in participating due to the disability
Parents and guardians		• Parents are acting on behalf of their child and therefore may not have an appreciation of population level effects (e.g. herd immunity with vaccines).
People who use health and/or social care services		
Carers	• Carers have direct experience of ill health, which the patient may not be able to communicate. • Carers are also affected by the person who is suffering from a disease/condition and impact's their lives.	
Members of the public		

Organisations who represent the interests of these individuals	• Organisations are likely to have a good understanding of patients' views and feelings, which may be hard to capture with a small number of patients • Organisations may be better placed to convey information to patients	

Initially we planned to sample members of patient organisations similarly to the Wave Two case study of rimonabant. For rimonabant, we recruited obese individuals through a patient organisation based at University College London called Weight Concern (http://www.weightconcern.org.uk/). This was the result of a clearly specified process:

1. Contacting a number of patient organisations identified through the PCWP, IAPO, web searches and people who work within the field of obesity (including referrals)
2. Identifying those who would be willing to participate and most strongly represent the patient population defined in our decision problem
3. Once we had identified the group Weight Concern, we were required us to complete a number of tasks:
    - Obtain ethical approval from the relevant institution
    - Agree on the payment amount
    - Agree on timelines (they do not like to overwhelm their members with too many surveys)
    - Receive approval once they review the survey and make amendments where necessary
    - Agree to circulate the results to all members in a newsletter type format

For natalizumab, we tried to adopt the same approach. However, we encountered several challenges: (1) difficulties identifying and contacting MS patient organisations based in the United Kingdom, (2) difficulties for patient organisations to collaborate with us as they either had challenges dedicating resources to advocate our research, and/or focused primarily on how to practically solve the immediate everyday concerns of patients. As a consequence, we discussed alternative methods of recruitment and modified our approach from contacting patient organisations and instead contacted an MS Clinic at Imperial College who agreed to participate and allow the PPI team to access registered patients currently receiving treatments. Another potential method we could have used is crowdsourcing which is currently used by the FDA, which uses large number of people to perform smaller tasks.[1] However, we chose not to explore this method of recruitment owing to its unfamiliarity and the time constraints involved.

Dr. Richard Nicholas aided us with our ethical application and provided guidance regarding different methods of sampling MS patients at Charing Cross and St. Mary's Hospital. Charing Cross Hospital has a dedicated clinic space for those receiving natalizumab (approximately 330 patients). Patients are there for two hours when they attend their appointment and receive treatment. The St. Mary's Hospital MS Clinic shares a waiting room space with other medical specialties (e.g. cardiology department) and a range of different MS treatments are offered to patients (approximately 3000). He discussed the advantages and disadvantages of recruiting from each location: e.g. at St. Mary's the patients are there for short time and quickly leave their appointment; if we publicly advertised and asked

[1] http://www.fdanews.com/articles/152443-fda-approves-first-trial-protocol-designed-via-8216-crowdsourcing-8217?v=preview

them to fill out questionnaires, many would not like to self-identify as having MS in a waiting room and would not be willing to approach researchers.

We decided that we would like to sample patients from both clinics as we would like to adopt a pragmatic approach and cover a variety of patients. Based on sample size considerations discussed within each of methodology sections, we aimed to recruit 500 individuals.

We consider the following:

- How should the patients be approached?
  - Introduction in a confidential manner by the physician or patient approaches us?
    - **Physician recommends our study in the patient appointment as we think this we result in better recruitment**
  - Does the physician hand out (a) information (person in next room is doing a study would you like to think about participating), (b) questionnaire (read, consent if you'd like and either place in box or go to person in next room)?
    - **Physician hands out information sheet**
  - Public stand/set up or separate room?
    - **We will sit in a separate room and hand out questionnaires for participants to fill in (complete then and there and place in a box). There is a minor query with this though, ethically perhaps patients should be given an option to post the questionnaire back as informed consent requires 24 hours in most cases**
- How many patients do we want?
  - There are a huge number, how do we determine which we'd like?
    - Which treatment they're on?
    - As many as possible?
    - As fast as possible?
    - Randomly pick days to show up at different clinics?
  - **We want a variety of patients (the broader it is, the more pragmatic it is) on different treatments who have had the condition for different durations, the number will be informed by the sample size calculation and time required to recruit**

3.4.2 Ethical approval

Although ethical approval had previously been sought and granted from the Imperial College Research Ethics Committee (ICREC Reference Numbers: 12_4_5 and 12_2_8), this form of ethical approval has the limitation that it does not provide approval for direct access patients or their data through the NHS, and it is not possible for any of the research to take place on NHS premises or use NHS facilities. For ethical approval to cover the sampling of patients through the NHS we had to apply through IRAS.

3.4.3 Questionnaire development

3.4.3.1 Questioning method

It is important to acknowledge that depending on which structured method of preference elicitation is used, there may be multiple ways to administer the required questions. For example focus groups, interviews, telephone calls, emails, paper questionnaires and online surveys could all potentially be used. However, the selection of the questioning method might depend on a variety of factors. Some factors discussed by the team were:

- Population characteristics: e.g. older populations may have less access the internet, the progression of some illnesses might mean that patients may not have the dexterity to use a pencil or computer mouse.
- Complexity of the questioning method: some elicitation methods may require the completion of a large number of complicated tasks. However, in some cases complex questions can be broken down and simplified and questions can be divided between the population sample.
- Dropout: for minimal dropout it is necessary for the method to be engaging; complex formats, difficult to answer questions and lack of engaging stimuli may result in fewer completed questionnaires.

Whichever preference elicitation and questioning method is selected, we considered it essential that participants sufficiently understand the tasks to provide valid answers.

3.4.3.2 Defining outcomes and clinical consequences

When eliciting weights from patients and the public, a certain amount of clinical knowledge about the disease and adverse events of drugs for disease treatment is assumed. However, for a successful elicitation process it is important for subjects to understand precisely what is meant by each benefit and risk criteria and also what the clinical consequences are for each of them. Information was compiled to define what is meant by each criterion in the benefit-risk tree and what the short and long term implications are for each of them.

We need to be clear: if we incorrectly describe the outcomes (particularly the unfavourable outcomes), the preferences we elicit may be biased. In the team we determined key criteria which had to be met, i.e.

- Naming the outcome
- Describing the outcome (i.e. listing symptoms)
- Stating how often the outcome is likely to occur
- Stating the duration the outcome is likely to last for
- Saying if the outcome can be treated or not

The descriptions of the outcomes were created by the team with specific input from clinicians and review by IAPO. There were debates regarding the content of the boxes as specific outcomes may be unclear.

We picked to present the definitions in a tabular format

  - The advantage is that it is easy to search for specific information and compare across the outcomes; although, we note a caveat, some people may have difficulties reading tables

Specific examples of difficulties:

- Seizures
  - hard to find a clear description of seizures in regulatory documents (e.g. SmPC); there are many different types of seizures

- difficulties associated with being exhaustive: we need concise descriptions suitable for patients which should focus on the most important aspects of treatment

Medical perspective is very necessary.

Care is required when describing risks

- EMA documents/slides: including undesirable effects requires caution, to prevent "undue alarm", "fear", and "inappropriate emphasis on side effects"
- Is it possible that if we design MCDAs, DCEs and AHPs we may cause "undue alarm", "fear" and "inappropriate emphasis" because we describe so many potential risks? Or do not provide information on the magnitude…?
- May introduce more questions and make the questionnaire longer

Concerns with using clinical measures: EDSS could be alienating…

3.4.3.3 Feedback to participants

We explicitly ask participants to provide us with their email address if they wish to know the results of the research. The results will be sent out in a newsletter format which will include: a recap of the study context/purpose/aim, the results of the study, and how the contribution of the participant has helped the field of patient and public involvement in the benefit-risk assessment of medicines. They will also be provided with an online link to the full report on the IMI PROTECT website should they wish to learn more in depth about the study.

3.4.4 Piloting the questionnaires

Ideally, the questionnaire should be rigorously piloted. Because of the size of our sample and difficulties accessing the patients prior to obtaining ethical approval, we chose alternative methods of piloting our work. They included asking the following groups represented in table XX for their thoughts on the comprehension and accessibility of the questions in addition to any comments and concerns. There are advantages and disadvantages to piloting the questionnaire which each group.

Group	Advantage(s)	Disadvantage(s)
Colleagues in WP5 who are members of the PPI team	Representation of a wide range of professional perspectives	Close to the material as involved in its development
Colleagues in WP5 who are not members of the PPI team	Extensive knowledge of benefit-risk assessment methodologies, includes people who analyse the content from a clinical, industry, academic, patient organisation perspective, easily available, large in number	missing the lay perspective, don't have the condition
external colleagues and acquaintances without MS	provide external perspective	don't have the condition so can't provide patient perspective
colleagues and acquaintances with MS	have the condition and can represent a patient perspective, easily accessible/convenient	missing the lay perspective
patient organization	capable of reading from patient perspective, developed similar materials before	patient organisation is not specifically related to MS although it does cover it

NB all of this is done on an informal basis, with feedback kept confidential and anonymous; anyone asked has the option to decline.

3.4.4.1 Feedback from pilot

General feedback

Examples of general feedback are shown below. Feedback specific to the AHP and DCE questionnaires is given in 3.5.6and 3.5.7 respectively.

General

- Requests for additional information (e.g. better descriptions of symptoms)
- Request for a reduction in amount of information provided
- Duration of estimated time required to complete the questionnaire should be stated
- Some words have specific clinical meanings and need to be used with care, even though patients will be completing the questionnaire (e.g. serious)
- General things: formatting, spelling errors, cut and paste errors
- Need to state if personal or population level perspective is desired (importance to society or person)
- Simplification of language, e.g. replace "quantify" with "measure"
- Confusion: favourable/unfavourable, effects/outcomes/measures, benefits/risks
    - benefit and risk because they're familiar words within most peoples lexicon; however, as Larry often says, it's not a word which has a set definition within people's minds
    - The discussion continues: benefits and risks, favourable and unfavourable effects, advantages and disadvantages of treatment:
        - We initially specified the terminology benefits and risks in our documentation
        - The terms favourable and unfavourable effects have been suggested previously (Phillips, EMA Benefit-Risk Project)
        - We replaced benefits and risks with favourable and unfavourable effects in the PPI documents
        - Following the changes, we discussed the terminology with a patient organisation and they strongly suggested to remove favourable and unfavourable effects and advocated benefits and risks
        - We then changed the terminology back to benefits and risks following the advice from the patient organisation
        - We will keep benefits and risks because we consider it to be appropriate with a non-technical audience (although we acknowledge the caveat that patients may have different definitions/concepts of benefits and risks in their head)
        - Something which happen on the "side" but they can be extremely central and important in the patient experience
- People with MS may have vision difficulties; is the font used appropriate, is it too small?
    - From our experience in reviewing the literature on visual data representation, there is not much written on this and guidance is hard to find. We acknowledge that:
        - Questions must be appropriately spaced on the page. Descriptions and questions must also appear together on the same page.
        - A bigger font, although desirable for the patients, may lead the questionnaires to having too many pages which may be daunting
- Sometimes there was conflict (e.g. more information required vs. too much information is given)
    - Sometimes substantial changes were requested

  - In these cases, decisions were made at the discretion of the PPI team; our decisions were founded on:
    - The volume of feedback requesting the change
    - Professional and personal perspective(s)
    - Limitations of what we could accomplish within the timelines
- It is extremely important to consider "Which treatments would YOU take?" versus "Which treatments should be available for a patient population?" It is also important to note that patients do not make decisions in isolation, their physicians guide them
  - A proposal was needed on the different ways of asking the question which reflect the regulatory and clinical decision context; a physician team member made a proposal
  - How is this addressed/reflected in the questionnaire?
  - Patients use themselves as a natural frame of reference rather than thinking at a population level perspective; team agreed to use proposed wording but with minor changes (i.e. to remove "his/her choice of" in the questionnaires)

3.5 Additional notes on multi-criteria methods

We completed some preliminary work prior to eliciting preferences.

Multiple criteria decision analysis methods require objective data, i.e. data on the performance of a treatment. It also requires subjective data, i.e. the value function and weight, which may vary according to stakeholder perspective. These two types of data are then combined to evaluate the overall value of an alternative. We investigated different methods of obtaining weights, i.e. AHP, swing-weighting, and MACBETH for the same alternatives, and criteria, while assuming a linear value function. This section presents a description of each of the methods we used to elicit preferences. Please note that for each method, participants were requested to provide their own values rather than take the perspective of a regulator making decisions on behalf of other patients with RRMS and the public.

3.5.1 Definitions

When examining how the three different MCDA benefit-risk methodologies can be used to elicit preferences from patients and the public, it was noted that the use of benefit-risk assessment terminology is often confusing and inconsistent. For example, different authors may use different words to refer to the same process; alternatively, different authors may also use the same word to refer to different processes. To prevent confusion, the following definitions provided in Table 3-9 were adopted for terms relating to the methodologies in this protocol (Belton and Stewart, 2001).

Table 3-9 Definitions of benefit-risk assessment terminology

Term	Definition
Alternatives a	The choice of treatment
Criteria i	The set of consequences which will be used to compare the alternatives
Performance = $z_i(a)$	The measure of performance of alternative according to criteria i
Partial value function = $v_i\big(z_i(a)\big) = v_i(a)$	Maps from the performance to a number called the value score.
Weight = w_i	Reflects the importance of criterion i
Value = $V(a) = \sum_i w_i v_i(a)$	The overall value of alternative a. In this protocol the benefit-risk score.

3.5.2 Linear and non-linear value functions

All three methods require use of a value function, which is a mathematical function based on subjective judgement that defines the relationship between data and utilities. AHP requires a value function based on the performance of each of the treatments under consideration, which can either be performed by constructing a matrix with multiple levels of outcome that cover the data ranges of the treatments under consideration, or directly comparing the exact data of the treatments within a matrix while blinding the names. MCDA and MACBETH provide the option of either using a linear value function, or constructing a piecewise monotonous non-linear value function.

In the natalizumab team there were two schools of thought:

1. To not ask patients for their value function preferences; assume that value function is linear, and use it in all subsequent analyses
2. To ask patients for their value function preferences; derive the value function, and use the derived value function in all subsequent analyses

Strong arguments were presented for both sides of the debate (Table 3-10).

Table 3-10 A Comparison of Linear versus non-linear value functions

Linear	Non linear
Ethical basis: one life cannot be worth more than another, each case is of equal importance	Logical basis: individual perception of risk probabilities is unlikely to be linear
Regulatory, societal level and public health perspective	Individual, personal level, and medical perspective
Good for proportions and percentages	Good for severities, and rare and serious adverse events that do not have a denominator (i.e. number of cases)
Less burden on researcher and participant: fewer preference elicitation questions required	Increased burden on researcher and participant: more preference elicitation questions required
In most cases a linear value function is the best approximation when the real shape of the value function is not known, i.e. it is not feasible, or too difficult to obtain a preference based value function	Ideal when it is both feasible and possible to obtain a preference based value function

In total there were three options available to the case study team:

1. Assume linear value functions for all benefit and risk outcomes
2. Elicit patient preferences to derive value functions for all benefit and risk outcomes
3. Use a mixture of linear, and preference based value functions, i.e. elicit for the benefit and risk outcomes that were considered to have had the greatest importance in the natalizumab marketing authorisation decision, and use linear value functions for the remaining benefit and risk outcomes

After extensive discussions, it was decided that a linear additive value function would be assumed for all methodologies (except where interaction terms were to be investigated as part of the DCE). The rationale for this was to reduce the amount of cognitive burden on the participants, and reduce the amount of time required for the completing questionnaires and decision conferences.

3.5.3 Levels and swings

With two of the methods (DCE and swing weighting) it is necessary to determine both an upper and lower data value for each continuously measured outcome in the value tree. These values are simply used as anchor points for preference elicitation, and do not have a direct impact on the results of the benefit-risk assessment as they are not intended to reflect the performance of any particular treatment. However, there are points to consider when selecting these values so as to minimise distortion in the elicitation process and ensure that the resulting preference weights are appropriate to the decision problem. The upper and lower levels for each outcome should be:

- broadly consistent with those seen in treatments in the real world;
- sufficient to cover the range of outcomes among the alternatives in the decision problem;
- easily communicable to and understandable by participants;
- chosen so that the utility of the swing from lower to upper is broadly comparable between outcomes (i.e. the utilities of the swings are reasonably close in order of magnitude terms)

After considerable discussion and piloting of the relevant questionnaires, the upper and lower values set out in Table 3-11 were adopted. To ensure a consistent approach, the same upper and lower levels were used for both the DCE and swing weighting methods, although different levels could also have been used.

Table 3-11 The upper and lower values used for preference elicitation for each continuously measured outcome. The relapse figures are 2-year rates; the binary outcomes are expressed in terms of the number of patients experiencing each event out of 1000 over a 2-year period.

Outcome	Lower value	Upper value
Relapse	1.5 relapses	2 relapses
Disability progression	100 out of 1000	250 out of 1000
Herpetic encephalitis	0 out of 1000	1 out of 1000
PML	0 out of 1000	3 out of 1000
Transaminases elevation	30 out of 1000	60 out of 1000
Seizures	0 out of 1000	40 out of 1000
Injection site reactions	0 out of 1000	250 out of 1000
Allergic reactions	0 out of 1000	500 out of 1000
Flu-like reactions	250 out of 1000	750 out of 1000
Serious allergic reactions	0 out of 1000	100 out of 1000
Depression	100 out of 1000	200 out of 1000

3.5.4 Swing weighting

The swing weighting methodology is performed in a group setting led by a facilitator. Firstly, an information letter, consent form, glossary and MCDA swing-weighting questionnaire will be distributed to those who wish to participate in the study. Next, two meetings will be scheduled according to the availability of those willing to participate. The first meeting will occur on an individual basis via telephone, and will be aimed at providing an overview of the project and to give instructions on how to complete the MCDA swing-weighting questionnaire. Each participant will be asked to complete the questionnaire between the two meetings. The second meeting will be a group decision conference. All of the participants meet together in person to present their individual answers to the questionnaires, and then come to a group consensus for the elicited weights.

Weights are elicited throughout the hierarchy of the value tree. Starting with each outcome measure, the relative ranks of a swing from the best to the worst of each outcome within each category are derived. Starting from where all outcome scores are at their worst, participants are asked to choose the outcome they would most want to move to the best score. They are then asked to rank the other outcomes in a similar way. This acts as a "thought stepping

stone" for then assigning preference weights on these outcomes. For this, the top ranked outcome is given a weight of 100 and marked on a visual analogue scale, and participants then place the other outcomes on the scale to reflect their relative importance.

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

3.5.5 MACBETH

MACBETH is a software-based decision-making methodology which aims to elicit and numerically represent qualitative value preferences to compare specific benefit and risk criteria. Similar to MCDA, an information letter, consent form, glossary and MACBETH questionnaire will be distributed to those who wish to participate in the study. Next, two meetings will be scheduled according to the availability of those willing to participate. The first meeting will occur on an individual basis via telephone, and will be aimed at providing an overview of the project and to give instructions on how to complete the MACBETH questionnaire. Each participant will be asked to complete the questionnaire between the two meetings. The second meeting will be a group decision conference. All of the participants meet together in person to present their individual answers to the questionnaires, and then come to a group consensus for the elicited weights, in this case using the MACBETH software tool to express their preferences in qualitative terms (in contrast to the quantitative approach used with MCDA).

3.5.6 AHP

Analytic hierarchy process (AHP) uses pairwise comparisons between criteria, where preferences are assigned using a numeric scale. These comparisons are then combined into a single scale of priorities, which are expressed as a single number for the best alternative, or by a vector of priorities that order alternatives proportionally. Whereas MACBETH performs constant consistency checks during the decision-making process, AHP only performs a consistency check once the decision-making process has ended. As it does not require amendments relating to consistency before progressing to the next stage, it was possible to format the AHP questions into a printed survey which was distributed to patients at MS clinics in London.

Feedback from pilot questionnaire:

- Important is not a good word to use, especially in the methods of administration section
    - Replace all with preferred/desirable
- Directionality was an issue: the wording and format was perhaps confusing and some people had to think about the meanings of "less", "more", "positive", negative"
- The form will be long and complicated for a patient to fill in on their own, it either needs to be administered by someone or simplified

3.5.7 DCE

It is recommended that a DCE should contain no more than seven attributes (outcomes); if there are a greater number then the cognitive load increases and participants may tend to ignore certain attributes when responding to questions.

In order to assess the DCE method, it was therefore decided to concentrate on a subset of the outcomes in the value tree and following discussion the following set of six attributes was adopted:

1. Relapses
2. Disability progression
3. PML
4. Mild allergic reactions
5. Serious allergic reactions
6. Depression

Within each choice question, participants will be asked to choose between hypothetical treatment options. It is assumed that participants choose the alternative with the highest utility. By assuming a particular form for the utility function (a function of the attribute levels plus a random component), the probability of making a choice can be fitted using regression models, with the regression coefficients corresponding to the impact of each attribute on overall utility.

The starting point is a simple additive linear utility function with no interactions (i.e. main effects only):

$$V = \beta_0 + \beta_1 * relapse + \beta_2 * disability\ progression + \cdots + \beta_6 * depression + e$$

Where V = utility, β_i= main effects regression coefficients and e=the random component of utility

If sufficient responses are received, an additional model will be run to test whether there are any significant two-way interactions between attribute preferences. This will indicate whether the assumption of preference independence underlying MCDA and similar methods holds in practice.

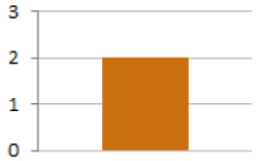
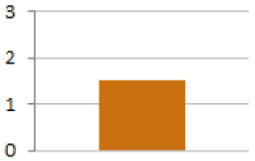
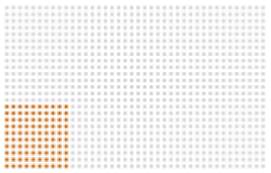
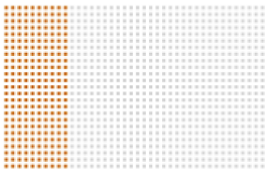
Although choice sets of three or more alternatives can be used in DCEs, a simple binary choice design between two alternatives was adopted in this case. This reduces cognitive strain on the participants and also facilitates the statistical analysis. Some DCEs (eg in market research applications) allow respondents to opt out of each question or indicate that they have no preference between the options in order to capture this aspect of real-world decision making. As we are simply using the method to elicit the underlying attitudes to treatment outcomes, and patients in the real world are generally not faced with stark choices like those in the DCE, we have not allowed opt-outs and instead force respondents to choose between the two alternatives presented.

A fractional factorial design was generated in the SAS software package, consisting of 32 choice sets. The design was then split into two 16-question surveys. Each participant would complete one survey, meaning two participants would be required for one full set of responses. The questionnaires would be distributed in alternating order to help ensure balance. Based on simple rules of thumb and efficiency statistics, it was judged that this design would provide the right balance between statistical power (estimates will become more precise with more choice sets) and brevity (participants may be put off by a long survey or become disengaged and lose focus). To further ease the potential burden on participants, it was agreed to prompt participants halfway through each survey (i.e. after 8 questions) and allow them to stop there if they so wish. Each 16-questionsurvey was therefore split into two blocks, and alternating versions of the survey (with the order of the blocks changing each time) were used for balance. An additional question was added to each 8-question block (giving 9 overall), aimed at testing whether the respondent understood the task and was engaging correctly; these validation questions had “correct answers” in that there was no utility trade-off between the alternatives as one was clearly of higher utility than the other. If a validation question was found to be answered incorrectly then the other 8 responses in that block would be discarded.

Within each choice question, the outcome levels were communicated in text and by use of visual representations. After discussion it was agreed to use bar graphs for the relapse rate and icon graphs (showing the number of affected patients out of a group of 1000) for the risks, as shown in Figure 4 below.

Figure 4 Extract from DCE questionnaire

Consider imaginary treatments A and B for relapsing-remitting multiple sclerosis.

Outcome *(measured over 2 years)*	**Treatment A**	**Treatment B**
Number of relapses per patient	2 relapses 	 1 to 2 relapses
Disability progression	100 patients out of 1000 	 250 patients out of 1000

The bullet points below show the specific feedback from the DCE pilot questionnaire:

- A number of changes have been made, including:
  - Introductory text and instructions have been edited/completed
  - Descriptions of scales have been added
- Number of relapses:
  - The scale ends at 2 relapses which may give the wrong impression to a patient that this number is the absolute maximum; this will be expanded to 3
  - Query: Do we need a graph?
    - We're aiming to try and have a balanced representation of benefits and risks; we have used visualisations for all of the risks and so we are aiming to use them for the benefits too
    - From work in the visual review and elsewhere, it would be difficult to try and find another easily interpretable graphical representation for this outcome
  - The label 1 to 2 relapses is a range; participants may think of either end of the range when evaluating their preferences
    - Statistically, this will be coded as 1.5 relapses in the analysis
    - Is it strange to have contrast between words presented in range but graphical representation is set to 1.5?
    - Verbally, it's difficult to describe fractions of relapses and this format was thought best for patient understanding
- Disability progression:
  - The labels 10% risk and 25% risk need to be reversed
  - Query: Background vs. part-to-whole information; the scale only goes to 25%
    - This will be changed to 100
  - An icon array might be a different way of presenting the information but it is questionable for this specific outcome
- Risks:
  - Colours used:
    - Brighter colours?
      - For people who are colour blind, contrast may be more important than colour

        - Bright colours (i.e. red in this case) might be more emotive rather than informative and influence choices
    - Using a magnifying glass for rare risks with not many squares shaded in
        - From a visual perspective this may not be desirable because it is not an impartial method of displaying evidence; the illustrator has made a subjective decision based on their opinion which may prejudice those viewing it

3.6 Results

The framework, completed as far as possible at this interim stage, is provided in Table 3-12.

The complete results, including those of the preference elicitation work, will be reported in due course.

Table 3-12 Preliminary application of the framework

Step	Point to consider	Application
Determine the purpose of PPI	What is the aim?	The aim is to elicit the values and preferences of patients and the public directly affected by RRMS regarding the benefits and risks of natalizumab, glatiramer acetate, and interferon beta-1a.
	During which stage(s) of the methodology is involvement required and/or desired?	Although it was desired to apply PPI to multiple stages of the decision-making process, this was infeasible due to the limited amount of resources which were available. Because of this, PPI was applied to the analysis stage of decision-making, and the quantitative framework benefit-risk assessment methodologies of MCDA swing-weighting, MACBETH and AHP were used.
	What is the desired level of involvement for each stage?	The desired level of involvement for the analysis stage was consultation. We wanted to test the feasibility of obtaining patient and public values and preferences on the benefits and risks of treatments for RRMS and compare the results.
Ethical approval	Is ethical approval required? Who needs to approve ethics application? How long is needed to obtain ethics approval?	Ethical approval was required. Imperial College Research Ethics Committee reviewed and approved the protocol. The Committee meets to review proposals once every two months and this was considered in the timeline of activities.
Address potential barriers and negative outcomes	What are the potential barriers to meaningful involvement?	We encountered multiple barriers: • Lay language: describing (a) the benefits and risks, and (b) the tasks required to be completed by the participant while still maintaining medical/technical precision was extremely difficult. • Format: In PPI literature, there are multiple different methods of involving people. However, the literature does not explicitly state which works best, and/or would be most appropriate for our setting, e.g. one-on-one with a facilitator, focus group with a facilitator, paper questionnaires, online questionnaires. • Timescales: PPI activities were not originally built into the WP5 timeline. Instead, they were later added as a "bolt-on" to the original activities, which meant that a limited extra time allowance could be allocated. The additional time required as a result of unanticipated barriers, e.g. difficulty recruiting, was hard to handle. • Participant burden: Each of the methodologies require a relatively large number of questions which place a cognitive and time burden on participants.
	What are the potential negative outcomes of involvement?	We anticipated that a potential barrier might be the perceived expectation that the results will be explicitly incorporated into the final decision.
	How can the barriers and negative outcomes be alleviated?	It is hard to know how recruitment, language, and format can be improved upon. This is because PPI in benefit-risk assessment is a relatively new concept and there is an insufficient evidence base to guide how these issues can be addressed. However, in future, we would recommend that PPI

Step	Point to consider	Application
		activities are carefully planned *a priori* to any project, and sufficient time is allocated for ethical approval processes and recruitment.
Training	Do participants require training and support? If so, how will this be addressed?	Is has been reported in the literature that the benefit-risk methodologies of MCDA swing-weighting, MACBETH, and AHP have been successfully completed with lay people. However, before involving participants, careful explanations of the processes need to be explained prior to involvement. We plan to have a physician present during the MCDA swing-weighting and MACBETH decision conferences who can inform participants if they have difficulties understanding the benefits and risks. For AHP, an email address will be provided at the start of the online survey for participants to contact if there is any confusion.
	Do researchers require training and support? If so, how will this be addressed?	Training and support was necessary due to the complexity of the methodologies. Training support for MCDA swing-weighting and MACBETH was provided by benefit-risk methodological experts; additional help was provided by staff at the MHRA for the development of the glossary.
Recruitment	Which group of participants will be used to represent patients and the public?	Participants who responded to the MS Society or MS-UK advertisement will be used to represent the views of patients and the public.
	What is the sample size required for the methodology?	There is no minimum sample size required for MCDA swing-weighting, MACBETH, or AHP.
	What are the methods of recruitment?	Emails introducing to our study were sent to several national MS organisations in the UK with a request for help to recruit. We followed this up with telephone calls. The MS Society and MS-UK offered to place advertisements on their website.
Design a statement of agreement between the researchers and participants	Is it possible to provide full disclosure of the benefit-risk methodology being studied and the role and anticipated value of patient and public involvement?	We did not provide full disclosure of all the tasks involved in the benefit-risk methodologies, e.g. statistical analysis. This is because we believed too many technical details and would be daunting and confusing to patients and the public. The role and anticipated value of patient and public involvement was carefully detailed to potential participants in the recruitment email and introduction to the study. It was hoped that by describing the value of involvement, individuals would have greater motivation to participate.
	What are the roles and responsibilities of both the researcher and the participant?	In this study, we believed the role of the researchers were: ○ To design a scientifically accurate study ○ To ensure the study went through an external ethical review process and adhere to ethical principles (e.g. confidentiality and anonymity) ○ To limit the burden placed upon participants ○ To describe the study and tasks in a manner easily comprehensible so that participants clearly understand what is asked of them

Step	Point to consider	Application
		In this study, we believed the role of the participant was: ○ To be engaged, and provide truthful responses
	Have the expected needs and contribution of both the researcher and the participant been defined?	This step was not completed. See commentary.
	Do the participants know all of the confidentiality, anonymity, drop-out, and acknowledgement policies?	For MCDA swing-weighting and MACBETH, information on confidentiality, anonymity and dropping-out were detailed in the information sheet prior to participation. For AHP, confidentiality and anonymity were detailed at the start of the questionnaire, and it was assumed that if people wished to drop out they would close the internet browser.
Patient involvement activities	What are the anticipated time scales for involvement activities?	Originally we hoped we could complete MCDA swing-weighting, MACBETH, and AHP in WP5 Wave Two. Unfortunately, there were significant delays to this due to issues with recruitment.
	What is the method of communication?	• AHP questionnaire: online survey • Participant information leaflets: email • Dates for decision conference: email and telephone • Decision conference: face to face • Dissemination of results: email
	What is the location of involvement activities (if applicable)?	The decision-conference for MCDA swing-weighting and MACBETH took place at Imperial College, London. It was held at a location which is wheelchair accessible. AHP took the format of an online questionnaire.
Finances	Are there finances in place to specifically support involvement?	There were no finances specifically in place to support involvement at Imperial College, London. These were later negotiated between PROTECT management and WP5 partners.
	Can participants receive adequate financial support for their expenses and contribution?	For MCDA swing-weighting and MACBETH, participants were reimbursed £80 to cover expenses for their travel and time. This is a standard amount for reimbursement per diem by patient organisations. For AHP, the decision was made not to provide financial support for participant contribution. This is because the level of burden placed upon the participant to complete an online questionnaire is low.
Reporting of outcomes	What were found to be the benefits of involvement?	The preference elicitation work is still in progress, and so this step cannot be completed.

Step	Point to consider	Application
	What were the negative outcomes of patient involvement?	• Recruitment of participants: We found it extremely difficult to recruit participants. This is because we could not approach patients directly within clinical settings, as research involving NHS patients, their data, or NHS premises or facilities requires IRAS or NRES ethical review. The time required for this type of approval would have exceeded the timeframe of PROTECT WP5. Instead, we approached multiple national MS organisations via email and telephone: although many saw the importance of our research, they were limited in the help which they could provide, e.g. they were willing to advertise on their website but could not help with more active methods of recruitment. The number of responses from interested participants as a result of the website advertisements was few.
Reporting of conflicting perspectives	Did conflicting perspectives or disagreements occur?	The preference elicitation work is still in progress, and so this step cannot be completed.
	At which stage of the process did they occur?	The preference elicitation work is still in progress, and so this step cannot be completed.
	Who did they occur between?	The preference elicitation work is still in progress, and so this step cannot be completed.
	What were the different perspectives?	The preference elicitation work is still in progress, and so this step cannot be completed.
	How were they resolved?	The preference elicitation work is still in progress, and so this step cannot be completed.
Communication	Is it necessary for participants to be periodically informed of the decision-making process as it progresses?	Yes, both at the start (participant information sheets), and at the end of the process (circulation of results).
	How will participants be informed about the results of their involvement?	Participants were informed about the results of their involvement via email.
	How will the contribution of participants be explicitly acknowledged?	We anticipated that the contribution of participants would be explicitly acknowledged, and they could opt for their names to be included in the report, while ensuring that the confidentiality of responses was maintained.
	How will the involvement process be reported to all stakeholders?	The process will be reported to (a) participants, and (b) to patient organisations.
	How will the overall impact of the decision on patients be evaluated?	In this study, we are examining the feasibility of eliciting preferences an authorisation decision is not being made. However, we will record evaluative feedback from the participants once the results have been circulated.

4 Discussion and Conclusions

4.1 Summary of completed and ongoing research

Much of the work has been completed but due to delays in obtaining ethical approval, the preference elicitation work is ongoing and this has also impacted the completion of related work elsewhere in the project. A summary of the completed and ongoing work is presented in below. Updated results will be provided in due course.

Table 4-1 Completed and ongoing research

Key area	Aim	Work completed	Work ongoing
Patient and public involvement and the benefit-risk assessment roadmap	To assess where and when in the benefit-risk assessment decision-making roadmap PPI is desirable and/or feasible.	• Examination of benefit-risk frameworks, methodologies and current practice • Assessment of where and when PPI can be applied in the BR roadmap • Development of a framework for applying PPI in BR assessment	• Testing the PPI framework in practice
Preference elicitation from patients and the public	To elicit weights from patients and the public regarding benefit and risk outcomes for the treatment of relapsing remitting multiple sclerosis (RRMS) using the benefit-risk methodologies of swing-weighting, MACBETH, DCE, and AHP.	• Case study has been refined with new benefit and risk data • Planning of elicitation studies and development of materials • DCE and AHP questionnaires distributed	• Swing weighting and MACBETH decision conferences to be held • DCE and AHP responses to be received • Results from all preference elicitation methods to be analysed
Visual representations	To test visual formats for the communication of different stages of benefit-risk assessment with patients and the public and healthcare professionals.	• Two surveys on preferences for type of visual displays for weight loss - one for lay people for themselves, and one for healthcare professionals for the people they treat or advise • Brief report on lay preferences completed	• Analysis of survey responses • Full technical report
Communicating the outputs of WP5 to participants, patients, and the public	To create a lay summary of the recommendations and PPI work.	• Website established • Initial materials developed	• Further online materials to be developed

A full discussion of the results, their significance and external validity will be provided when analysis has been completed.

4.2 Recommendations for future research

There are several directions future research for the assessment of patient and public involvement in the assessment of benefit and risk of medicines could proceed.

Firstly, we have assessed a collection of common methodologies for preference elicitation. However, there are many other methodologies than could also be assessed in the context of benefit-risk assessment. Specifically, methods which are well established in other disciplines, may not behave in the same way in a medical context where very mild and very severe adverse events may need to be compared. Methods also need to be applicable to patients whose disease may impact the reliability of preference elicitation tools.

Secondly, there are relatively few example of applying preference elicitation methods to patients, so further case studies exploring the applicability of these methods in a variety of indications and patient populations would be beneficial.

Thirdly, in the context of drug development, research could be perform to assess which stages of benefit-risk evaluation during the drug development process are most suitable for patient involvement. This may be dependent on the context of the drug and patient population, as there may be situations where patient involvement is central to the regulatory decision making, an others where it adds little value.

Finally, regulators make decisions at a population level, whereas preference election methods are performed on a specific group of patients. To understand the impact of this on regulatory decision making, we need to understand how individual patient characteristics impact preferences, and which patients may be the most appropriate to include in a preference election study.

5 Appendices

5.1 Information sheet

Imperial College London

School of Public Health
Imperial College London

LG33A, Medical School Building
Norfolk Place, London W2 1PG
Tel: +44 (0) 2075941738

kimberley.hockley08@imperial.ac.uk
www.imperial.ac.uk

Information Sheet

Study: Eliciting Patient Preferences on the Benefits and Risks of Treatments for Relapsing Remitting Multiple Sclerosis

We would like to invite you to participate in this original research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

Research Aims

We are interested in patient preferences on the benefits and risks of treatments for relapsing remitting multiple sclerosis. We are specifically interested in how we can ask patients for their preferences using special types of formal questioning methods and how these different methods compare against one another. We hope that this study will provide helpful information on how to involve patients in decisions made by medical doctors, policymakers, and regulators and make sure that their preferences have been considered. At the conclusion of the study, we will provide you with a newsletter summarising the main findings.

Who Have We Asked to Participate?

We have invited patients with relapsing remitting multiple sclerosis from St. Mary's Hospital Paddington and Charing Cross Hospital Hammersmith to take part in the project.

What Will You Be Asked to Do?

We will ask you to complete a paper questionnaire and post it back to us using the envelope which is provided. You will be asked to answer some questions about your age, gender, when you first experienced symptoms of multiple sclerosis and how you compare and value different treatment outcomes. In the questionnaire, we will also ask if you would like to participate in a focus group to further discuss your responses using a different type of questioning method.

When and Where Will the Study Take Place?

We ask you to complete the paper questionnaire contained within this study pack at a location of your choice. Then, we kindly ask you to return it using the envelope provided. If you do decide to participate, we kindly request that you return the questionnaire to us as soon as possible within two weeks of your appointment. The focus groups will take place at St. Mary's Hospital or Charing Cross Hospital.

What happens if I would like to take part in the focus group?

If you do decide that you would like to take part in a focus group, we request that you provide us with your name, email and telephone number in the questionnaire. We will then contact a number of those who have indicated that they are willing to participate to find out if they have any additional questions and check their availability. Before the focus group, we will require signed consent. For your records, a copy of the consent form is contained within this study pack.

What does participating in a focus group require?

To participate in a focus group you must be willing to do the following things:

1. Provide us with your name and telephone number or email address in the survey.
2. Receive a briefing about the study, which will describe the study and discuss any further questions you may have about the study and participating.
3. Travel to the St. Mary's Hospital Paddington or Charing Cross Hospital, Hammersmith.
4. Provide informed consent and participate in a group discussion session to talk about how you value different treatment outcomes.
5. Be willing for the discussion to be recorded. The recording will be anonymised, remain confidential and be securely stored.

How Long Will the Study Last?

The paper questionnaire will require approximately 30 to 45 minutes for you to complete. The focus group will last for approximately six hours.

Will You Compensate Me for My Time?

Unfortunately, we cannot compensate you for the time required to complete the paper questionnaire. If you participate in a focus group, to thank you we will offer you refreshments and lunch on the day, and we will also reimburse your travel expenses and additionally provide £75 in compensation for your time. You are under no obligation to participate in a focus group if you do decide to complete the questionnaire.

Are There Any Risks Involved in Participating?

The risks involved in participating are minimal. If there are questions that you find distressing or intrusive, you are free to not answer those questions or to withdraw from participating. You can withdraw from participation at any time without penalty.

Are There Any Benefits Involved in Participating?

The results of the study will increase the evidence base for policymakers and regulatory decision-makers to involve patients and the public in decision-making. Regulators are likely to benefit from more effective methods of public and patient involvement, resulting in more democratic, ethical and informed decision-making.

How Will We Maintain Your Privacy and Confidentiality?

Everything you tell us will remain completely confidential within the limits of the law. We will give you an identification number to replace any information we have in the data file that identifies your name, and your address or email address. The information that you give us will be completely anonymised and linked only to the numerical ID. Your information will be stored on secure computers in locked offices and in locked file cabinets. At the conclusion of the study, these data may be made available in totally anonymous form to other researchers outside of the project team. This means that researchers outside our project team will have access to the answers you provided in the questionnaire and the focus group if you participate. However, they will have no means of identifying you.

What will happen with the results of this study?

At the conclusion of the project, if you provide us with your email or postal address, we will send you a newsletter describing the major findings. Results of broad scientific interest will also be published in scientific journals and presented in meetings and conferences.

Who is Organising and Funding the Research?

This study is being organised by a team of researchers at Imperial College London. The Principal Investigator is Dr Kimberley Hockley, of Imperial College London. The study is being funded by the Innovative Medicines Initiative.

What If I Have Questions about the Project?

Please contact Kimberley Hockley by email (Kimberley.hockley08@imperial.ac.uk) or via telephone (020 759 41738).

Who Has Reviewed this Study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by National Research Ethics Service (NRES) Committee South Central - Oxford B (Reference number 14/SC/1442).

What If I Have a Complaint?

If you are unhappy about the way this research is being conducted and would like to make a complaint, please contact the Patient Advice and Liaison Service (PALS) (http://www.nhs.uk). If your complaint has not been resolved, please contact the Imperial College Joint Research Compliance Office (jrco@imperial.ac.uk).

5.2 Consent form

CONSENT FORM FOR FOCUS GROUP

Title of Project: Eliciting Patient Preferences on the Benefits and Risks of Treatments for Relapsing Remitting Multiple Sclerosis

Name of Researcher: Kimberley Hockley

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated 25th November 2014 (version 1.1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that data collected during the study will be looked at by individuals from Imperial College London and regulatory authorities. ☐

4. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent.	Date	Signature