Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

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

Patient and Public Involvement Team

Study Protocol (26th November 2014 V1.2)

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

Authors	Organisation
Kimberley Hockley (team leader) Deborah Ashby Subhakanta Das Christine Hallgreen Shahrul Mt-Isa Ed Waddingham	Imperial College London
Georgy Genov	European Medicines Agency
Yasemin Dil Jo Groves Rebecca Johnson Alison Lightbourne Jeremiah Mwangi Rachel Seal-Jones	International Alliance of Patient Organisations
Craig Allen Andrew Thomson	Medicines and Healthcare products Regulatory 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 Medicine 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

Sponsor

Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Regulatory Compliance at:

Joint Research Compliance Office
Imperial College London & Imperial College Healthcare NHS Trust
5th Floor, Lab Block
Charing Cross Hospital
Fulham Palace Road
London
W6 8RF
Tel: 0203 311 0212
Fax: 0203 311 0203

Contents

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 Patient 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

Background and rationale

Introduction

The work in this study 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 plan to explore and address these concerns with the study described in this protocol which aims to evaluate patient preferences for some of the favourable and unfavourable outcomes of natalizumab—a treatment for relapsing remitting multiple sclerosis.

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. 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.

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.

Involvement of patients and the public in this research

Representatives from the International Alliance of Patient 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, they 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 questionnaires have been piloted with individuals with multiple sclerosis for feedback (e.g. to assess comprehension and ease of completing tasks).

Study 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 benefits and risks 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 is to elicit patient preferences on the benefits and risks of RRMS treatments using different methods of elicitation proposed for use in benefit-risk assessment. This will be achieved through the following objectives:

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

It is important to note that processes described and conclusions drawn from the planned study presented herein relate solely to the testing of methodologies and representations for the evaluation of benefit and risk of medicines; our work neither replaces nor is intended to replace or comment on any regulatory decisions made by national regulatory agencies, nor the European Medicines Agency.

Study design

Here we outline our approach when developing our study to elicit preferences from patients and the public. Key steps are described here:

1. Specification of the decision problem
2. Selecting which method(s) of eliciting preferences to use
3. Evaluating the process of preference elicitation

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.

(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.

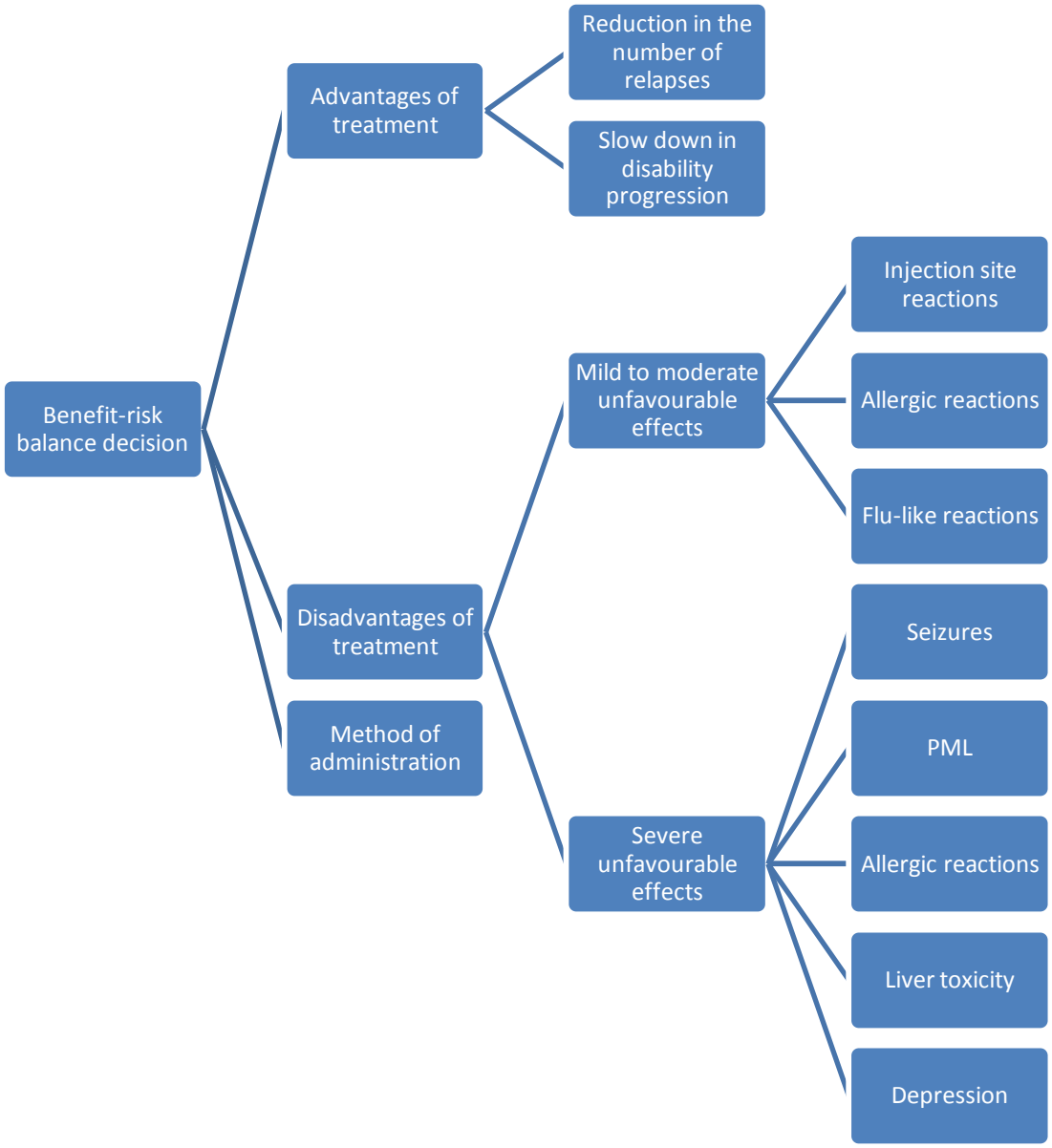
(d) Favourable and unfavourable outcomes

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 0.1.

Figure 0.1 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.

Selecting which method(s) of eliciting preferences 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.

Analytic Hierarchy Process

AHP is another decision-making method that can take into account multiple risks and benefits simultaneously.

(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 0.1). 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 0.2), a total of 12 (i.e. 3 + 3 x 3) pairwise comparisons need to be made.

Table 0.1 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
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 0.2), 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.

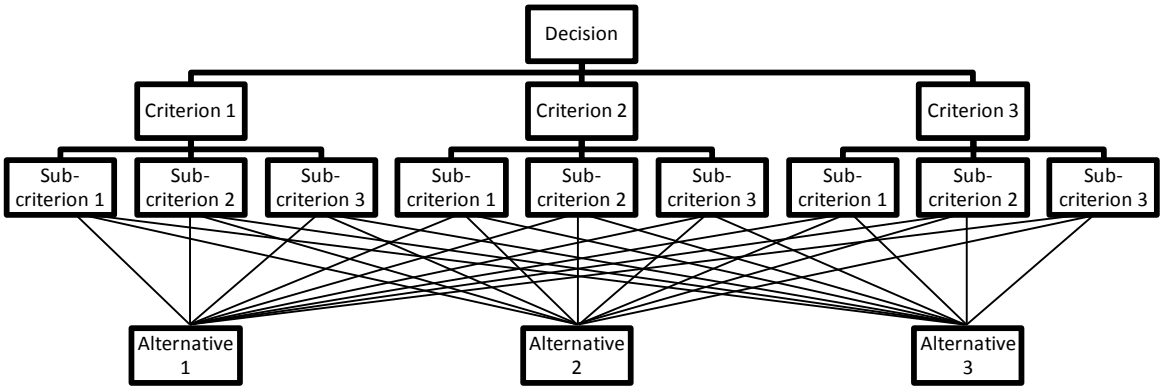


Figure 0.2 Illustration of an analytical hierarchical process including 3 criteria, each with 3 sub-criteria, and 3 alternatives

(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.

The analytic hierarchy process questionnaire can be found in **Error! Reference source not found.**

Discrete Choice Experiments

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 $(x = 1,2,\ldots,K)$.

The discrete choice experiment questionnaires can be found in **Error! Reference source not found.**

Swing-weighting

Swing-weighting is a type of MCDA. 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).

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.

Evaluating the process of involvement

The perspective of the participant

For each of the methodologies, participants will be asked how they feel 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 will rate how strongly they agree or disagree 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."

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

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 0.2). We plan to complete this as the study progresses.

Table 0.2 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?

	Step	Points to consider
	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?
	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?

	Step	Points to consider
		How will the overall impact of the decision on patients be evaluated?

Participant selection and recruitment

Screening and enrolment

We plan to recruit a sample of patients receiving treatments for relapsing remitting multiple sclerosis from Charing Cross and St. Mary's Hospital, London, England. Treating clinicians will screen potential participants during routine appointments. Those who meet the inclusion criteria will be provided with a study pack (a stamped addressed envelope containing an information sheet, questionnaire, and consent form for a focus group).

Patients will take the study pack home and can 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 will be included in the study pack for returning the questionnaire. A subset of participants expressing interest in participating to the focus group will then be asked to participate. The right of the individual to refuse participation will be respected and they did not have to provide reason or justification. Once participants enter the study, they will be free to withdraw at any time without giving reasons and without prejudicing further treatment.

Inclusion and exclusion criteria

Potential participants will be included in our study if they are:

- 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 will be excluded if they are:

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

Participant information sheet

The participant information sheet is presented in Appendix XX.

Consent

The right of the individual to refuse participation will be respected and they do not have to provide reason or justification. All participants once entered the study will be free to withdraw at any time without giving reasons and without prejudicing further treatment. A completed, returned questionnaire will be judged as implied consent. For the focus groups, written consent is required. An information sheet and consent will be provided in the study pack for participants to review and consider.

Statistical analyses

Sample size and power considerations

Discrete choice experiment

No established formal sample size formula exists for discrete choice experiments, as the answer is highly dependent on the experimental design and the quantities one wishes to estimate, which can vary enormously between DCEs. Various rules of thumb for the sample size do exist, however, of which the most commonly cited is that of Johnson and Orme and states that n > 500c/ta where t is the number of choice sets (=at least 8 per participant), a is the number of alternatives per choice set (=2), and c is the highest number of levels of any attribute (=2) or, if two-way interactions are included, the highest product of levels of two attributes (=4). (Reference: Orme, B. 2010 Getting Started with Conjoint Analysis: Strategies for Product

Design and Pricing Research. Second Edition, Madison, Wis.: Research Publishers LLC). This gives n > 63 for the primary analysis (main effects) and n > 125 for the secondary analysis (interactions). If each participant completes more than 8 choice sets, the required sample size would be reduced.

We have also performed some exploratory modelling of the beta parameters, their standard errors and the choice probabilities in the statistical model (see A62). Whilst these are approximate calculations based on informative prior assumptions and should not be regarded as formal sample size calculations, they suggest that a sample size of 200-300 should be sufficient to detect (at 95% significance) attributes with substantial beta coefficients.

Analytic hierarchy process

The standard method of analysis for the analytic hierarchy process does not provide any measure of statistical significance and so there is no minimum sample size requirement. An alternative regression-based approach may be used but no formal sample size calculation has been performed; instead the same sample size as the DCE group will be used so that we can see how the methods compare.

Swing-weighting and MACBETH

There are no sample size requirements for swing-weighting and MACBETH focus groups. We will be aiming to include between 4 to 6 participants in each focus group. This number allows for a variety of opinions to be included in the focus group whilst also enabling each participant to express their opinions within the timeframe of the focus group.

Data analysis

Discrete choice experiment

The DCE results will be analysed using a binomial logistic choice model frequently seen in the DCE literature. For the primary analysis, each participant's utility is assumed to be a linear combination of the attributes Xi (i =1 to 6) represented in the DCE, plus a random component:

$$\text{Utility} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon$$

The β_i coefficients are the quantities to be estimated. They represent the amount of influence each attribute has on the overall utility of treatment. The binomial logistic model assumes that the random terms ϵ are drawn from independent, identically distributed extreme value type I distributions. From this the likelihood of the observed responses to the choice tasks can be derived, assuming each respondent chooses whichever alternative has the highest utility, and estimates of the β_i are then obtained by maximum likelihood estimation.

Analytic hierarchy process

The priority of each criterion in AHP can be derived using matrix algebra, where these are the principal right eigenvector of a matrix comprising of each pairwise comparison included in a hierarchy. An inconsistency factor is also calculated which provides a measure of inconsistency across comparisons made within the matrix for each hierarchy. As a rule of thumb, an inconsistency of 0.1 or less is considered suitable. Priorities assigned to each criterion will be summarised across participants. Measures of inconsistency across participants will also be summarised.

Swing-weighting

Swing-weighting will be carried out using the computer software HiView 3. The software guides the facilitator across the various steps of the framework. The first steps of the swing-weighting process will be carried out by the team before the focus groups: identifying the decision content and identifying the alternative and the criteria. The weighting of the criteria will be performed during the focus group. The software will then provide the weights assigned to each criterion.

MACBETH

MACBETH will be implemented using the computer software HiView 3. In this process, each criterion is compared to others in a pairwise fashion similarly as with AHP. However, this method deals with inconsistencies between various comparisons, where the software will alert the facilitator that an inconsistency is present, and will offer options on how it can be addressed. The software will then provide weights assigned to each criterion.

Regulatory issues

Indemnity

Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study.

Sponsor

Imperial College London will act as the main Sponsor for this study. Delegated responsibilities will be assigned to the NHS trusts taking part in this study.

Funding

Funding for Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) was received from the Innovative Medicine 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.

The Alliance of Patient Organisations will be reimbursing each participant in the decision conference £80 for their transport and time. Funding for reimbursement is also provided as stated above.

Audits

The study may be subject to inspection and audit by Imperial College London under their remit as sponsor and other regulatory bodies to ensure adherence to GCP and the NHS Research Governance Framework for Health and Social Care (2nd edition).

Ethical issues

This study will be submitted to the Research Ethics Committee through the Integrated Research Application System.

We do not expect our study to raise significant ethical issues. However, several issues were considered:

- Burden on participants: The methodologies require patients' time and can potentially involve a large number of questions. However, keeping in mind the serious nature of relapsing remitting multiple sclerosis, the number of questions will be kept to a methodological minimum to reduce the amount of burden placed on the participants. Participants will also be fairly reimbursed for their time; advice regarding this has been provided by the International Alliance of Patient Organizations.
- Anonymity and confidentiality: The amount of personal data collected will be kept to a minimum. To ensure confidentiality, person-identifiable information will be anonymized and not be used unless it is absolutely necessary. Access to confidential information will be operated on a strict need-to-know basis by nominated people within the research team. Those who are able to access the data will be made aware of their responsibilities and understand and comply with the law.
- Informed consent: All participants will be required to provide informed consent. A completed questionnaire will be assumed as a proxy for consent. Once entered, they will be free to withdraw at any time without giving reasons and without prejudicing further treatment.
- Patients may believe their participation or their responses could negatively impact current treatments. It will be stated that participation to the study and the choices made on the questionnaire will not affect or alter current care. Patients will be approached by their treating clinician in a sensitive and appropriate manner. Upset, angry, or distressed patients will be screened out of the study. Patient will take the study pack home and will be able to return the questionnaire by post.

- Questionnaires may make participants aware of unfavourable effects related to treatments which they were not aware of previously. They will have an opportunity to discuss this with their treating clinician.
- Participants may be worried about the financial burden of attending a decision conference. The International Alliance of Patient Organisations will be reimbursing each participant in the decision conference reimbursement £80 for their transport and time.

Confidentiality

Identifiable patient data will not be accessed for identifying potential participants to this study. Only participants expressing their interest in participating to further activities (i.e. focus groups for swing-weighting and MACBETH) detailed in this protocol and who provide their contact details will be invited to participate in these activities. Patient identifiable information collected in this study will only be accessible on a strict need-to-know basis by nominated people within the research team.

Timelines

Recruitment will begin in December 2014 and the study will end April 2014 when the results of the study will be circulated to participants.