BENEFIT-RISK ASSESSMENT

Summary Report of the Workshop on

The Development of a Model for Benefit-Risk Assessment of Medicines Based on Multi-Criteria Decision Analysis

organised by the
CMR International Institute for Regulatory Science
at the London School of Economics
29-30 March 2004

Stuart Walker
Margaret Cone

April 2004

CMR International Institute for Regulatory Science

Workshop on

THE DEVELOPMENT OF A MODEL FOR BENEFIT-RISK ASSESSMENT OF MEDICINES BASED ON MULTI-CRITERIA DECISION ANALYSIS

29-30 March 2004, London School of Economics

SUMMARY REPORT

Overview

Determining the benefit-risk balance of a medicine is one of the most important steps in its development, review and post-approval re-assessment but there are currently no well-established, validated models for this essential task. The objective of this interactive Workshop convened by the CMR International Institute for Regulatory Science was to examine a proposal for the development of a new model for benefit-risk assessment using Multi-Criteria Decision Analysis (MCDA). This is a method of looking at complex problems, of breaking the problem into more manageable pieces in order to allow data and judgments to be brought to bear on different aspect, and then of reassembling the pieces to present a coherent overall picture for decision makers. The purpose of this tool is to serve as an aid to thinking and decision making, but not to take the decision.

The Institute Workshop brought together senior experts from regulatory agencies and the pharmaceutical industry involved in the development, assessment and post-approval safety assessment of medicines for a preliminary evaluation of the model. (See participants list). The one-and-a-half day meeting was conducted as an interactive working session with the assistance of a facilitator, Professor Larry Philips, an expert in the techniques of 'Decision Conferencing' as well as MCDA. The proposed benefit-risk model was discussed and tested using a case study based on actual, but anonymised data.

Although time allowed only a relatively simple example to be discussed, participants were unanimous in agreeing that the methodology had great potential for addressing benefit-risk assessments and there was a strong wish to carry the project forward. From the industry point of view it was perceived as a means not only to enhance in-house decision-making processes about the viability of R&D projects but also to anticipate potential problems before an application is submitted for agency review. The regulators were keen to test the model further using retrospective and prospective data with a view to gauging its value in improving the consistency and transparency of the decision making process.

It was strongly recommended that the CMR International Institute should follow up this initiative with further Workshops, as soon as possible. In the meantime both regulatory and industry participants expressed the wish to study the technique further and learn more about the MCDA methodology.

Format of the Workshop

The main focus of the meeting was the live, interactive demonstration of the MCDA model, using the case study, in which delegates participated in the role of the assessors and the model was developed and tested 'on screen' using customised Hiview software[1]. A full report on the methodology and outcome of the case study is given in *Annex 1*. Before starting work on the case study Filip Mussen (Research Fellow developing the model as part of his doctoral thesis) presented an overview of the criteria and current methods for benefit-risk analysis (*Annex 2*) and Professor Larry Philips introduced the principles of MCDA. There was a wide ranging discussion on the concepts of the benefits of a medicine, in terms of efficacy, effectiveness and implications for the individual and public health as well as the assessment of risk and the differences in risk perception between the scientific and regulatory community and the public (see *Annex 3*).

The MCDA Model

The model discussed by the group and developed from the case study is presented here in broad outline only. (For more detailed information, please refer to the technical *Annexe 1*). The case study related to a hypothetical new recombinant tumour necrosis factor α (TNFα) receptor inhibitor indicated for the treatment of active rheumatoid arthritis resistant to other treatments, including methotrexate. Data on outcomes and adverse event reports were given for clinical trials on the drug alone against placebo and on the drug with concomitant use of methotrexate.

The first step in the MCDA model is to identify the ***Options*** for the possible outcome of the benefit-risk assessment. There might normally be only two options: that benefits outweigh risks (the product is the 'option' for treatment) or that risks outweigh benefits (the comparator or placebo is the option). In the specific case study, however, three outcome options were identified: the product, the product plus methotrexate and the placebo.

The second step is to identify the ***Criteria*** to be taken into account in determining the outcome ('option') for the assessment and to group these criteria under the two main headings of 'benefit' and 'risk'. In the case study, the criteria for benefit included results of the clinical studies measured as the percentage of patients achieving specific endpoints for symptomatic relief and slowing of disease progression. The criteria for risk included not only the incidence of adverse events and drug-related reactions, but also unobserved but potential risks based on knowledge of factors including related products, mechanism of action and likelihood of immunosuppression.

The software converts the criteria into a diagram described as the 'value tree' which gives a hierarchical structure for the factors contributing to the benefit-risk balance (see Annex 5). In a series of steps, assisted by the computer software, each criteria is assigned a ***Score*** and each score is given a ***Weight*** according to its relative importance to the benefit-risk decision. Weighted scores are calculated at each level in the hierarchy which enables an overall weighted score to be calculated for each of the options.

The process of 'scoring' is based predominantly on measurable data such as the clinical trial endpoints and incidence of adverse events, measured as percentages. Other scores, for example the risk of potential but unobserved reactions must, of necessity, be more subjective but the methodology, nonetheless, allows a numerical 'score' to be assigned such imprecise criteria.

The process of 'weighting' the scores is where experience and judgement are built into the methodology. The assignment of weight to a score is normally based on a combination of factors on which a value judgement is made, for example the scale of difference between the results for a drug and placebo in achieving symptomatic relief and the relative importance of demonstrating symptomatic relief over showing that the disease process has slowed.

Sensitivity testing and results

The 'subjective' element of scoring and weighting raised some concerns among participants, although it was accepted that, by their very nature, determinations of benefit-risk could never be based solely on objective evaluations of hard data. An important feature of the MCDA model and software, however, is the ability to carry out sensitivity analyses on the results by varying any of the weights and scores to assess the impact on the overall benefit-risk balance. Sensitivity analyses were demonstrated for some of the scoring and weighting decisions from the case study, where there had been differences of opinion amongst those present. Participants were surprised to note that, in several cases where the differences had appeared major, the impact on the overall outcome was scarcely significant.

The outcome for the particular case study and MCDA model tested at the Workshop was that the option with the optimal benefit-to-risk balance was the drug with concomitant methotrexate treatment. Participants were impressed to be informed that this was the 'real life' outcome for the assessment of the TNFα receptor inhibitor, Remicade® (inflixmab).

What was gained from the Workshop

Participants agreed that the workshop had been an important initial step towards developing a model with significant potential benefits for both companies and regulators. Multi-Criteria Decision Analysis, techniques in combination with the software, provide possibilities for a dynamic model that can be developed to and modified to meet a variety of scenarios encountered in drug development and evaluation.

The advantages of the MCDA technique that were identified during the meeting, included possibilities for:

- Enhancing the consistency and objectivity and transparency of the decision-making process for benefit-risk assessments by providing a structured and systematic approach and a 'paper trail' for tracking the process and providing greater accountability;
- Reviewing the consistency of regulatory decisions on marketing authorisation applications (MAA) in order to learn from past experience;
- Achieving a better understanding and more rational explanations of why different agencies reach different conclusions on the basis of the same data;
- Providing a training tool for both agency and industry staff involved in the development and assessment of new products;
- Allowing industry to test the benefit-risk data for new products before submitting an application, in order to identify areas where data may need to be strengthened or clarified;
- Carrying out more balanced and objective benefit-risk re-assessments in post-authorisation situations where there is a tendency to focus primarily on adverse event reporting.

(Further details of the discussion points are given in *Annex 3*)

The next steps

Participants urged the CMR International Institute to convene other interactive workshops using 'decision conferencing' techniques to explore further the possibilities of the MCDA methodology.

Acknowledging that the participation at this Workshop was predominantly European, Professor Walker proposed that priority should be given to a meeting in the USA in order to see if similar results and acceptance of the model would be achieved with FDA participation and more US company experts.

It was also suggested that more time should be allowed for subsequent workshops so that there could be industry and regulatory break-out sessions in order to see how different viewpoints impact on the scoring and weighting for benefit and risk.

Both industry and agency participants were keen to ensure that the impetus was not lost, in the meantime and wished to test the model further within their own environment.

Following the meeting possibilities for taking the project further in the EU were discussed with the EMEA.

It is proposed that the CMR International Institute for Regulatory Science should act as the primary contact point for companies and regulatory agencies that wish to follow-up this project and receive further information and contacts for developing the use of the MCDA model for benefit risk assessment.

Contact: Margaret Cone, Director of Regulatory Science, CMR International Institute for Regulatory Science

E-mail mcone@cmr.org, Telephone: +44 (0)1372 846100

References

1 Multi-criteria decision analysis was implemented in this project with the use of Hiview software, which was developed at the London School of Economics and is marketed by Catalyze Limited. Further information can be obtained at www.catalyze.co.uk.

Annexes

Annex 1 The MCDA model for Benefit-Risk Assessment of Product X, by Professor Larry Phillips

Annex 2 Overview of the criteria and current methods for benefit-risk analysis: Extracts from the presentation by Filip Mussen

Annex 3 Points from the Workshop discussions

Authors

Report prepared by Margaret Cone and Stuart Walker, CMR International. The contributions by Larry Phillips and Filip Mussen are gratefully acknowledged

WORKSHOP PARTICIPANTS

Facilitator

Professor Larry Phillips	Professor of Decision Analysis, London School of Economics

Delegates

Professor Gunnar Alván	Director General, Medical Products Agency, Sweden
Dr Daniel Brasseur	Chairman, Committee for Proprietary Medicinal Products (CPMP), EMEA
Dr Graham Burton	Senior Vice President, Celgene Corporation, USA
Dr. Janice Bush	Vice President, Safety Strategy and Liaison, Johnson & Johnson, USA
Ms Margaret Cone	Director of Regulatory Science, CMR International Institute for Regulatory Science
Dr Moira Daniels	Director, Global Regulatory Information and Intelligence, AstraZeneca Pharmaceuticals, UK
Dr Bryan Garber	Health Risk Management, Health Canada
Dr. Paul Huckle	Senior Vice President, GlaxoSmithKline R&D, UK.
Dr. David Jefferys	Head of Devices Sector, Medicines and Health Products, Regulatory Agency (MHRA), UK.
Dr. Patrick Le Courtois	Head of Unit for the Pre-authorisation Evaluation of Medicines for Human Use, EMEA
Dr Neil McAuslane	Chief Scientific Officer, CMR International Ltd
Mr. Filip Mussen	Director, Regulatory Affairs Europe, Merck Research Laboratories, Belgium
Dr. Sam Salek	Senior Lecturer, Department of Clinical Pharmacology, University of Wales, Cardiff
Dr. Tomas P Salmonson	CPMP Representative, Medical Products Agency, Sweden.
Dr. Valerie Simmons	Advisor, Global Product Safety, Lilly Research Laboratories, U.K.
Professor Stuart Walker	President and Founder, CMR International Limited, U.K.
Dr Beat Von Graffenreid	Division of Biotechnology Medicines, Swissmedic, Switzerland

Workshop Organisation

Workshop organised by: Professor Stuart Walker and Filip Mussen

Report prepared by Margaret Cone

CMR INTERNATIONAL INSTITUTE FOR REGULATORY SCIENCE

The CMR International Institute for Regulatory Science has been set up as a not-for-profit division of the Centre for Medicines Research International Ltd in order to continue its work in the regulatory and policy arena, and to maintain the well established links that the Centre has with regulatory authorities around the world. The Institute operates autonomously, with its own dedicated management, and funding that is provided by income from a membership scheme. The Institute for Regulatory Science has a distinct agenda dealing with regulatory affairs and their scientific basis, which is supported by an independent Advisory Board of regulatory experts.

CMR International, Institute for Regulatory Science,
Novellus Court, 61 South Street Epsom, Surrey KT18 7PX, UK
Tel: +44 (0) 1372 846100, Fax: +44 (0) 1372 846101
E-mail: institute@cmr.org

CMR International Institute for Regulatory Science

Workshop on

THE DEVELOPMENT OF A MODEL FOR BENEFIT-RISK ASSESSMENT OF MEDICINES BASED ON MULTI-CRITERIA DECISION ANALYSIS

29-30 March 2004, London School of Economics

ANNEX 1

THE MCDA MODEL FOR BENEFIT-RISK ASSESSMENT OF PRODUCT X

This Annex reports details of the model developed during the workshop. The report's structure follows the steps for carrying out a multi-criteria decision analysis outlined in (Dodgson, Spackman, Pearman, & Phillips, 2000), shown in Table 1.

Table 1. A summary of the MCDA process. Reproduced from the *MCA Manual*, p. 50.

Applying MCDA: Detailed steps
1. Establish the decision context. 1.1 Establish aims of the MCDA; identify decision makers and other key players. 1.2 Design the socio-technical system for conducting the MCDA. 1.3 Consider the context of the appraisal.
2. Identify the options to be appraised.
3. Identify objectives and criteria. 3.1 Identify criteria for assessing the consequences of each option. 3.2 Organise the criteria by clustering them under high-level and lower-level objectives in a hierarchy.
4. 'Scoring'. Assess the expected performance of each option against the criteria. Then assess the value associated with the consequences of each option for each criterion. 4.1 Describe the consequences of the options. 4.2 Score the options on the criteria. 4.3 Check the consistency of the scores on each criterion.
5. 'Weighting'. Assign weights for each of the criteria to reflect their relative importance to the decision.
6. Combine the weights and scores for each option to derive an overall value. 6.1 Calculate overall weighted scores at each level in the hierarchy. 6.2 Calculate overall weighted scores.
7. Examine the results.
8. Sensitivity analysis. 8.1 Conduct a sensitivity analysis: do other preferences or weights affect the overall ordering of the options? 8.2 Look at the advantage and disadvantages of selected options, and compare pairs of options. 8.3 Create possible new options that might be better than those originally considered. 8.4 Repeat the above steps until a 'requisite' model is obtained.

Establish context

Workshop participants were presented with a written case study which introduced Product X, a recombinant necrosis factor α (TNFα) receptor inhibitor. The proposed indication is for the treatment of active rheumatoid arthritis in adults when the response to disease-modifying anti-rheumatic drugs, including methotrexate, has been inadequate. It is the second TNF receptor inhibitor that will be marketed. A full preclinical and clinical program has been conducted and is described below. Product X is formulated as a solution for subcutaneous injection, and for the purpose of this case study it is assumed that there are no issues with the quality of the product. The proposed dose is 25 mg subcutaneous injection twice weekly.

The case study included information about pre-clinical data, clinical pharmacology and a phase IIb dose-ranging study. Results from two phase III trials showed, for the first study, the superior efficacy of Product X compared to placebo, and for the second study, the superiority of Product X plus methotrexate compared to placebo plus methotrexate. Safety studies presented in the case study revealed adverse reactions, some serious.

Identify options

Initially, two options were identified, Product X and placebo. However, when the group recognised that the second study added methotrexate to these two options, it became clear that three options were of concern:

1. Comparator: the placebo for Study 1 and placebo plus methotrexate for Study 2.
2. Product X
3. Product X plus Methotrexate

Identify objectives and criteria

The group agreed objectives common to all medicines, to maximise benefits while minimising risks. For this study those objectives were decomposed into concrete benefit criteria associated with study 1 and study 2, and risk criteria of specific adverse events and potential adverse effects. The structuring of objectives and criteria into a hierarchy is shown in Figure 1.

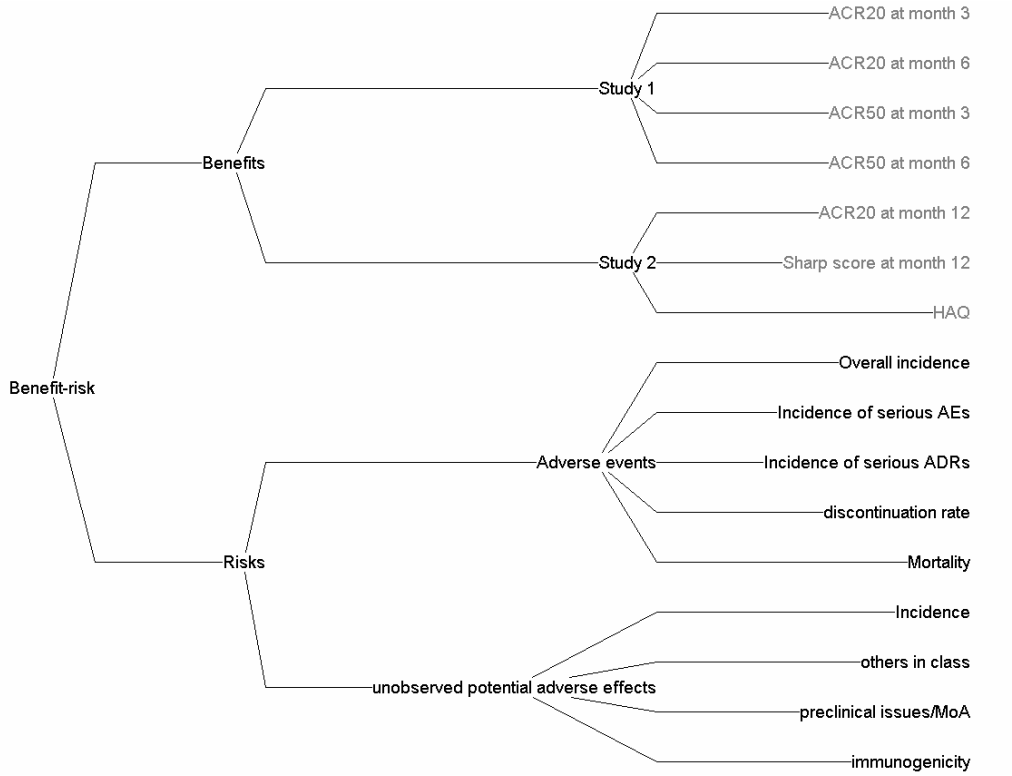


Figure 1: The hierarchy showing benefit and risk criteria.

ACR: refers to the American College of Rheumatology response criterion, a composite endpoint which requires at least 20% (50%) improvement in tender joint count and swollen joint count, plus 20% (50%) improvement in at least three of the following five items: patient assessment of pain, patient's global assessment of disease activity, physician's global assessment of disease activity, patient's assessment of physical function using the disability domain of the **HAQ** (Health Assessment Questionnaire) index, and acute phase reactant value, erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP). The time indicated by the criterion refers to the elapsed time after the start of treatment at which the ACR response was obtained.

Sharp score: refers to the total van der Heijde-modified Sharp score.

Overall incidence: Overall incidence of adverse reactions, in per cent.

Incidence of serious AEs: Incidence of serious adverse events, in per cent.

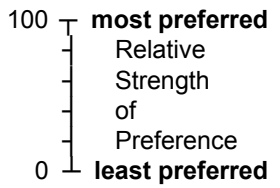
Incidence of serious ADRs: Incidence of serious drug-related adverse reactions, in per cent.

Discontinuation rate: Due to drug-related adverse reactions, in per cent.

Unobserved potential adverse effects: This cluster of four criteria attempts to capture current judgement about possible adverse effects that were not observed in the studies but might show up after launch.

Scoring the options

The purpose of scoring in MCDA is to achieve for each criterion a scale extending from 0 to 100, with those two points on the scale clearly defined. They can be defined independently of the options under consideration, or with reference to the least and most preferred options, as was done for this study. Scoring is the process of assigning numbers to the options on the criterion under consideration. The scale can be thought of as representing relative value or 'strength of preference'. Since it is an interval scale, like Celsius or Fahrenheit, whose zero points and units of measurement are arbitrarily chosen, it is important to appreciate that only differences in value can be compared on such scales. For example, if options A, B and C have been scored at 100, 80 and 0, respectively, then the difference in preference value between B and C is four times as big as the difference in value between A and B. It is wrong to suggest that option B, scoring 80, is 80% as good as option A, scoring 100, or that option D, scoring 40, is half as good as option B. It is ratios of differences in the scores that can be compared, not ratios of the scores themselves.



In this study, two approaches were used, direct and indirect scoring. Direct scoring required the group to identify the most and least preferred options on a given criterion, assign these scores of 100 and 0, then to score the remaining options so that differences in the scores reflected differences in preference. This approach was taken for all four criteria under the unobserved potential adverse effects node. Indirect scoring was applied to all the other criteria. Available data for these criteria provided inputs to the model, and the software converted these inputs to 0 – 100 scales, assigning 100 to the most preferred option and 0 to the least preferred. In the case of benefits, the translation was direct: higher numbers (e.g., higher ACR scores) converted to higher preference values. For the adverse events criteria (e.g., discontinuation rate), the conversion was inverse: lower input scores mapped to higher preference values. Thus, direct scoring enabled preference judgements to be input to the model in the absence of hard data, whereas indirect scoring used available data, which was linearly converted to preference scores.

The greyed-out criteria in Figure 1 indicate that data were not available for all options. For study 1, data were available only for Product X and the comparator, and for study 2 for Product X plus methotrexate and the comparator. As the absent data could contribute no preference value, the software assumed preference values of zero on these criteria. In effect, this meant that the absence of data was scored identically to the least preferred option, always the comparator, which also was given a score of zero because it was least preferred. Strictly speaking, this was incorrect, for the comparator was always associated with at least some value. For the purposes of this demonstration exercise, this didn't matter: the small difference between least value and no value was small. But for totally realistic applications it would be better to use fixed rather than relative scales, which would enable zero to be defined as meaning no value.

Weighting the criteria

Weighting in MCDA is the process of ensuring the equality of a unit of preference value on all the 0 to 100 scales. The scoring process results in a relative scale for each criterion, but the value difference between 0 and 100 may be different for each scale, as a Celsius degree represents a different unit of temperature from a Fahrenheit degree. The process of equating the units of value was accomplished by asking participants to compare the swings in preference from 0 to 100 on all the scales. This is a process of identifying the options associated with 0 and with 100 on a particular scale, then asking the group how big the difference is between those options and how much they care about that difference, as compared to 0-100 differences on other criteria. The process is called 'swing-weighting'[1].

It is this step that is perhaps the most misunderstood in MCDA, for weights are often thought to reflect the absolute importance of the criteria. Not so. The following example was given in the workshop. If you were to purchase a new car, would you consider price to be important to your decision? Most people answered, 'yes'. Since you can't consider all possible cars on the market, imagine that you construct a short list that includes just five cars. Suppose the difference in cost between the least and most expensive is £200. Now is price an important consideration in your decision? Most participants said 'no'. But if the difference in price were £2,000, then many said that would be more important. Unless, we pointed out, you are very wealthy, in which case a difference of £2,000 might not have much impact. The point of this example is two-fold: first, a criterion's weight depends on the range of difference in the input data, and secondly, on how much you care about that difference. Inevitably, that has to be a judgement. Thus, balancing risks and benefits requires judgement; it is inescapable, whether MCDA is used or not. While objective data may be available to establish the size of the difference between least and most preferred options on a criterion, the assessment of how much that difference matters requires, inevitably, an act of judgement. MCDA makes those judgements explicit.

Weights in MCDA represent trade-offs. Once the weights are established, they show how much an increase on one criterion is equal to an increase on another. An increase of 9° Fahrenheit is equal to an increase of 5° Celsius. Some methods for assessing weights rely on making judgements of trade-offs, or by comparing the best on criterion A and the worst on B with the worst on A and the best on B, but this latter method doesn't work when those hypothetical combinations are physically impossible. The more general technique, swing-weighting, is generally the preferred approach; it was used here. First, within a cluster of criteria, the largest swing was identified. It was given a weight of 100. Then all other swings on criteria within the cluster were compared to 100 and assigned appropriate weights. Thus, if the swing on another criterion was judged to provide half the swing in preference value, then that criterion was given a weight of 50. Weights are ratio scale numbers, since they compare differences, so there need not be a zero-weighted criterion.

For example, in study 2, the group judged that the difference between the comparator and Product X plus methotrexate was the largest difference they cared about for the Sharp score at month 12. Compared to that, the HAQ difference was judged to be nearly as important, so was assigned a weight of 95. The ACR at month 12 was judged to be the smallest swing in preference, and was given a weight of 60. Thus, the weights summed to 255, giving the total of the increments in value between least and most preferred options for study 2.

Weights for the study 1 criteria were assigned in a similar way, with the largest difference that mattered being that between Product X and the comparator for ACR50 at month six. That criterion was given a weight of 100, and the other three criteria assigned weights compared to that one, giving 50-80-60-100, respectively, a total of 290.

The next step was to compare the swing weight of 100 for study 1 with the swing weight of 100 for study 2. The group judged the comparator to Product X plus methotrexate difference in the Sharp score at month 12 to matter more than the comparator to Product X difference for ACR50 at month six. The relative weights for these two criteria were assigned as 100 and 40, respectively. The software then reduced the weights on all four study 1 criteria by 40 percent to ensure consistency. As a result of this equating of units of preference value across all seven benefit criteria, the total weight for study 1 became 116 (40% of 290) to study 2's 255. In short, study 2 contributes more than twice the overall value of study 1, in terms of distinguishing the product from the comparator.

A similar approach was used for the risk criteria. Finally, relative weights for Benefits and Risks were determined by comparing the one benefit criterion given a weight of 100 (Sharp score at 12 months) with the one risk criterion weighted 100 (incidence of serious ADRs). The group judged the former

[1] A brief introduction to swing weighting is given in Section 6.2.10, pp. 62-3, of the MCA Manual.

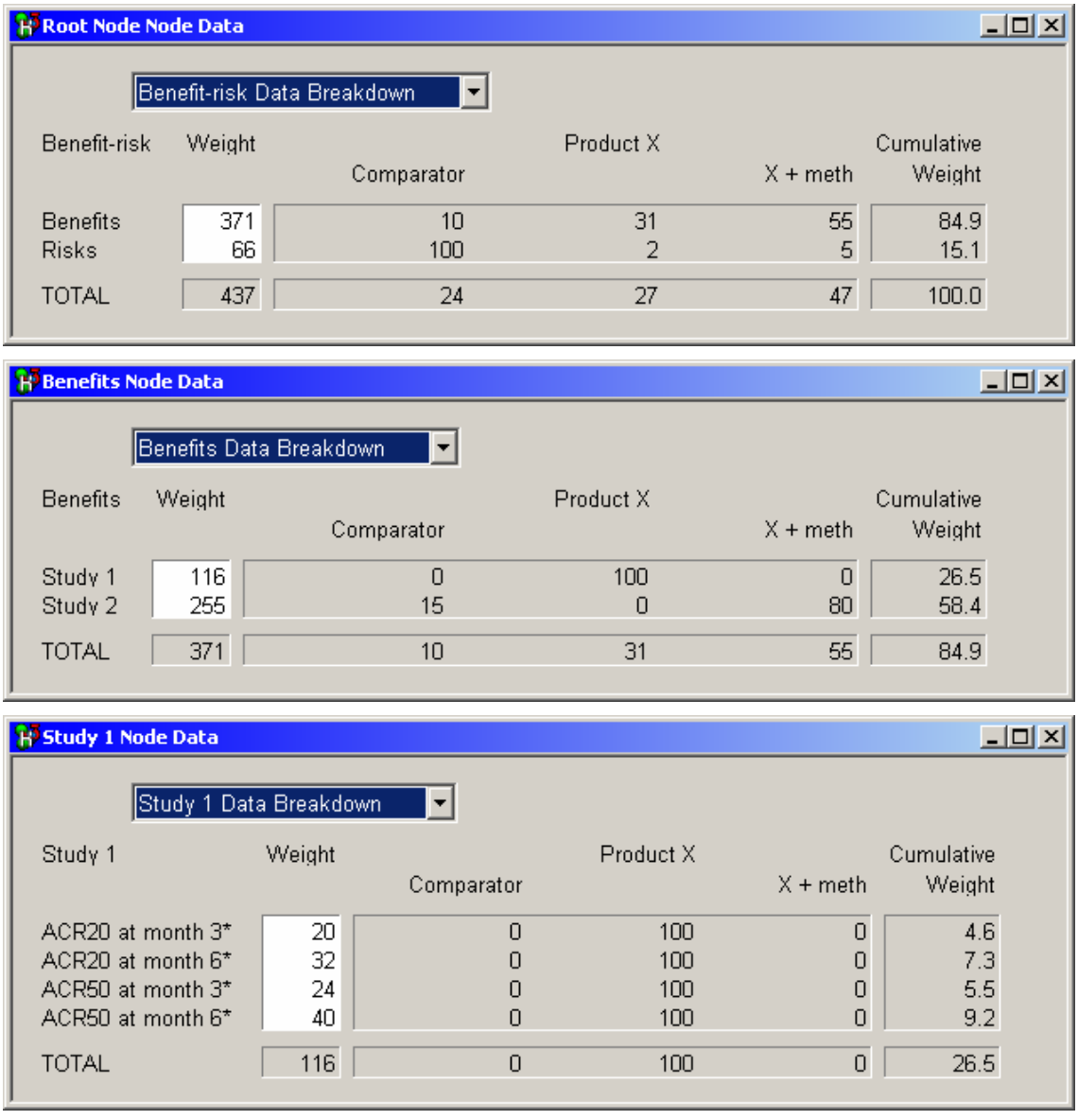
difference to be more important, so it was assigned a weight of 100, while the incidence of serious ADRs was give a weight of 20.

Although the explanation of weighting seems complex, in practice it was taken one step at a time, never requiring the group to compare more than two scales. The result is a set of very concrete comparisons, which while not necessarily easy to make, are considerably easier to do than making a holistic judgement of the relative importance of benefits and risks. The final result of this weighting process, a weight of 371 on benefits and 66 on risks, nearly a 6:1 ratio between the two clusters of criteria, would have been even more difficult for the group to judge, let alone justify. Rationale for the concrete comparisons of two criteria at a time could have been captured and written down, providing an audit trail for the final weights.

Combine weights and scores

The computer software does this. It normalises the weights at each node (making them add to 1.0 while preserving their ratios) and then multiplies the preference values for each option by the weights on the respective criteria. This process is repeated up through the value tree, giving a final set of weighted preference values.

The final preference values and weights for all the criteria and higher-level nodes are shown in Figure 3.



Root Node Node Data

Benefit-risk Data Breakdown

Benefit-risk	Weight	Comparator	Product X	X + meth	Cumulative Weight
Benefits	371	10	31	55	84.9
Risks	66	100	2	5	15.1
TOTAL	437	24	27	47	100.0

Benefits Node Data

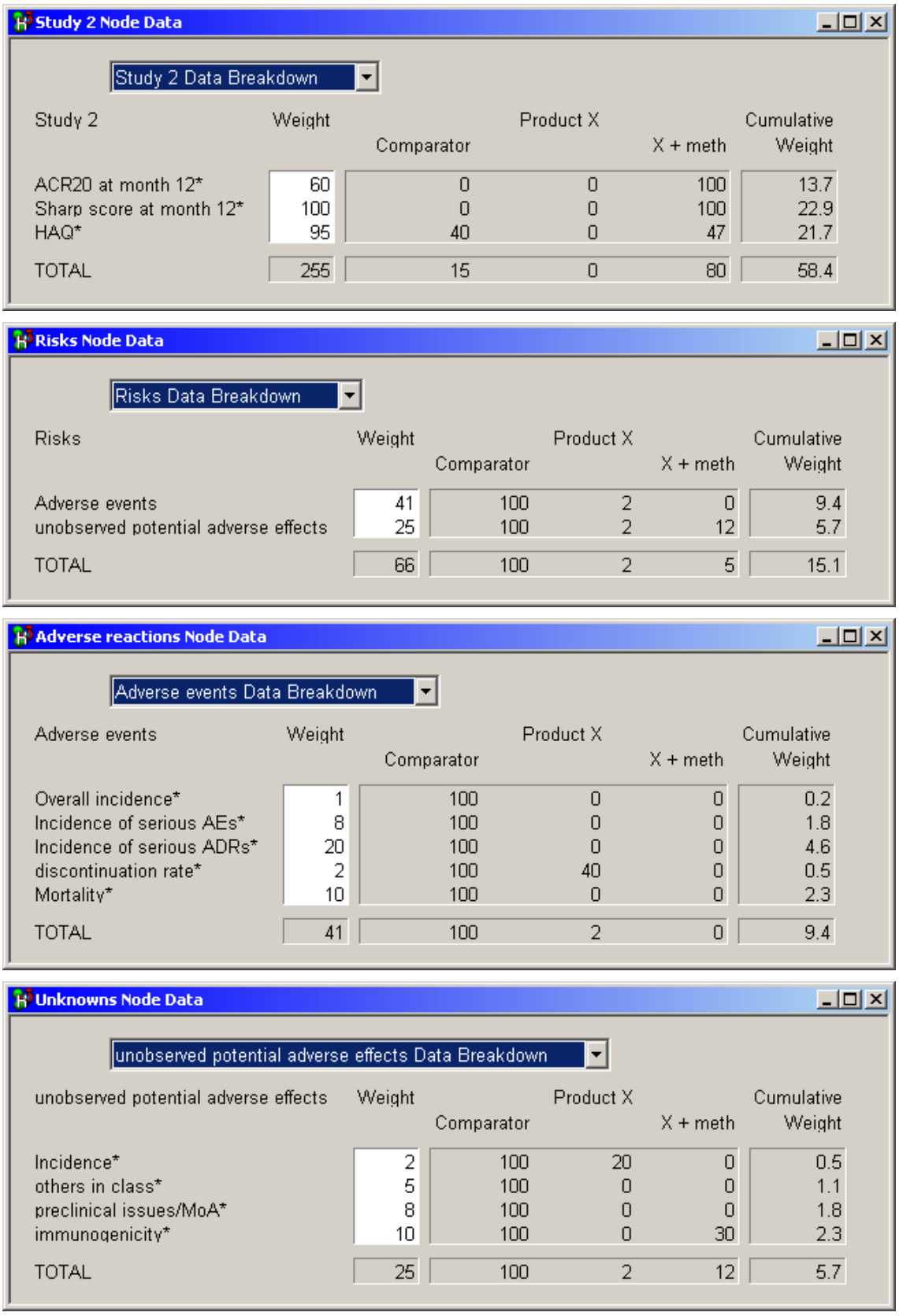
Benefits Data Breakdown

Benefits	Weight	Comparator	Product X	X + meth	Cumulative Weight
Study 1	116	0	100	0	26.5
Study 2	255	15	0	80	58.4
TOTAL	371	10	31	55	84.9

Study 1 Node Data

Study 1 Data Breakdown

Study 1	Weight	Comparator	Product X	X + meth	Cumulative Weight
ACR20 at month 3*	20	0	100	0	4.6
ACR20 at month 6*	32	0	100	0	7.3
ACR50 at month 3*	24	0	100	0	5.5
ACR50 at month 6*	40	0	100	0	9.2
TOTAL	116	0	100	0	26.5



Study 2 Node Data

Study 2 Data Breakdown

Study 2	Weight	Product X		Cumulative
		Comparator	X + meth	Weight
ACR20 at month 12*	60	0	100	13.7
Sharp score at month 12*	100	0	100	22.9
HAQ*	95	40	47	21.7
TOTAL	255	15	80	58.4

Risks Node Data

Risks Data Breakdown

Risks	Weight	Product X		Cumulative
		Comparator	X + meth	Weight
Adverse events	41	100	2	0	9.4
unobserved potential adverse effects	25	100	2	12	5.7
TOTAL	66	100	2	5	15.1

Adverse reactions Node Data

Adverse events Data Breakdown

Adverse events	Weight	Product X		Cumulative
		Comparator	X + meth	Weight
Overall incidence*	1	100	0	0	0.2
Incidence of serious AEs*	8	100	0	0	1.8
Incidence of serious ADRs*	20	100	0	0	4.6
discontinuation rate*	2	100	40	0	0.5
Mortality*	10	100	0	0	2.3
TOTAL	41	100	2	0	9.4

Unknowns Node Data

unobserved potential adverse effects Data Breakdown

unobserved potential adverse effects	Weight	Product X		Cumulative
		Comparator	X + meth	Weight
Incidence*	2	100	20	0	0.5
others in class*	5	100	0	0	1.1
preclinical issues/MoA*	8	100	0	0	1.8
immunogenicity*	10	100	0	30	2.3
TOTAL	25	100	2	12	5.7

Figure 3: Preference values and weights for the Product X benefit/risk model.

Examine results

The overall result is shown in the top matrix of Figure 3. Product X plus methotrexate scores 47 compared to Product X alone at 27. Product X scores little better than the comparator's 24, though for different reasons: some benefit and no risk for the comparator, compared to higher benefit and highest risk for Product X alone. Note, too, that Product X plus methotrexate is higher in benefits and lower in risk than Product X alone. These results can be seen graphically in Figure 4.

Figure 4: Overall results for the Product X benefit/risk model. Note that more green indicates more benefits, but because less risk is preferred to more, and preference is shown here, more red means less risk. Obviously, the comparator is least risky, Product X plus methotrexate is next, and Product X by itself is most risky.

It is clear that Product X plus methotrexate is overall most preferred. It is both more beneficial and less risky than Product X alone. The relative benefits considerably outweigh the relative risks, but this is largely due to the relatively higher weight placed by the group on benefits than risks.

Another way to see these results is to examine the overall weighted preference values versus the overall weighted risk preference, Figure 5. The comparator is lowest in risk (a high preference value) but lowest in overall benefits. Option 2, Product X alone, is highest in risk and moderate in benefits. Clearly, Product X plus methotrexate is slightly lower than Option 1 in risk and much higher in benefits.

Several participants found this to be an unexpected result. One participant, who knew the compound on which this disguised case study was based, pointed out that Product X plus

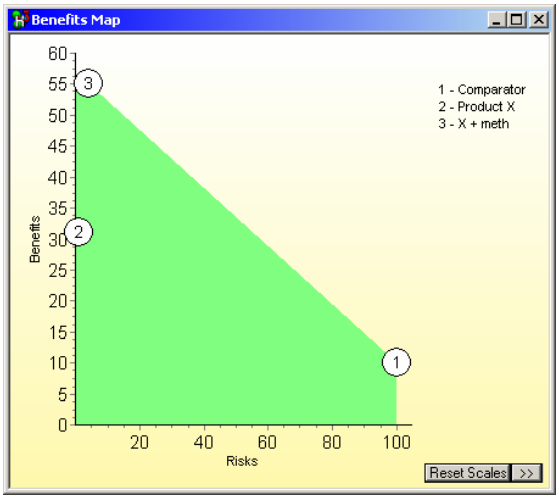


Figure 5: Benefits versus preference for risk for the Product X benefit/risk model.

methotrexate was eventually recognised as the best option, lending a degree of support to the MCDA model the group had constructed in rather a brief time.

Sensitivity analyses

Since the final result appears to be sensitive to the relative weight on risks and benefits, a sensitivity analysis showed the extent to which the final result was determined by these weights. This is shown in Figure 6.

The vertical red line indicates the relative weight on Risk, 15%, leaving 85% on benefits. At that weight, Option 3 is most preferred. That 15% has to be more than doubled, leaving less than 70% on benefits, for the most preferred option to become the Placebo. (The green shading only indicates the transition in overall preference.) Thus, a wide range of disagreement on these relative weights can be tolerated while those disagreeing can agree about the best option.

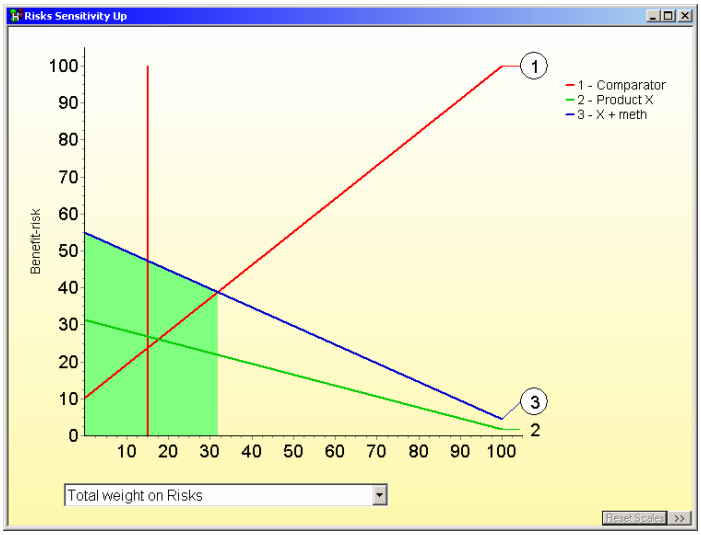


Figure 6: Sensitivity analysis on Risks.

Another sensitivity analysis showed the overall results as the weight on Adverse events was increased. Figure 7 shows that the current weight on this node was about 9%. That weight would have to be trebled before the Placebo would be preferred. Again, this shows that a decision about the benefit/risk trade-off does not depend on agreement about the relative weights between these two objectives.

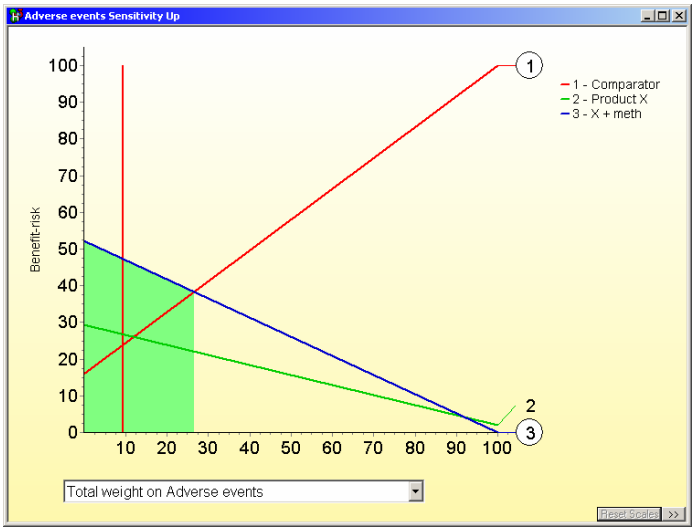


Figure 7: Sensitivity analysis on Adverse events.

In both these sensitivity analyses, as the weight on risk increases, the overall preference value of the placebo increases while the overall values of the other two options decreases, as would be expected.

Many more sensitivity analyses were performed to see if Product X plus methotrexate could be dislodged from its best position. It proved to be very robust to many different viewpoints about the relative importance of criteria and nodes. This is not, of course, always the case. Situations will arise in which one or more weights will make a difference. In this case, the MCDA model will not resolve those differences because they are essentially judgements. However, the sensitivity analyses are still useful in identifying areas in which more data might be required in order to resolve those differences in judgement. More importantly, the sensitivity analyses show that differences of opinion on many weights do not need to be resolved, so that more debate and discussion is unnecessary.

Sorts

These are another kind of sensitivity analysis, often useful in diagnosing why a particular option is better or worse, or in comparing two options. Although the feature was not explored during the decision conference, an example is shown here in Figure 8. Product X plus methotrexate is compared to the placebo.

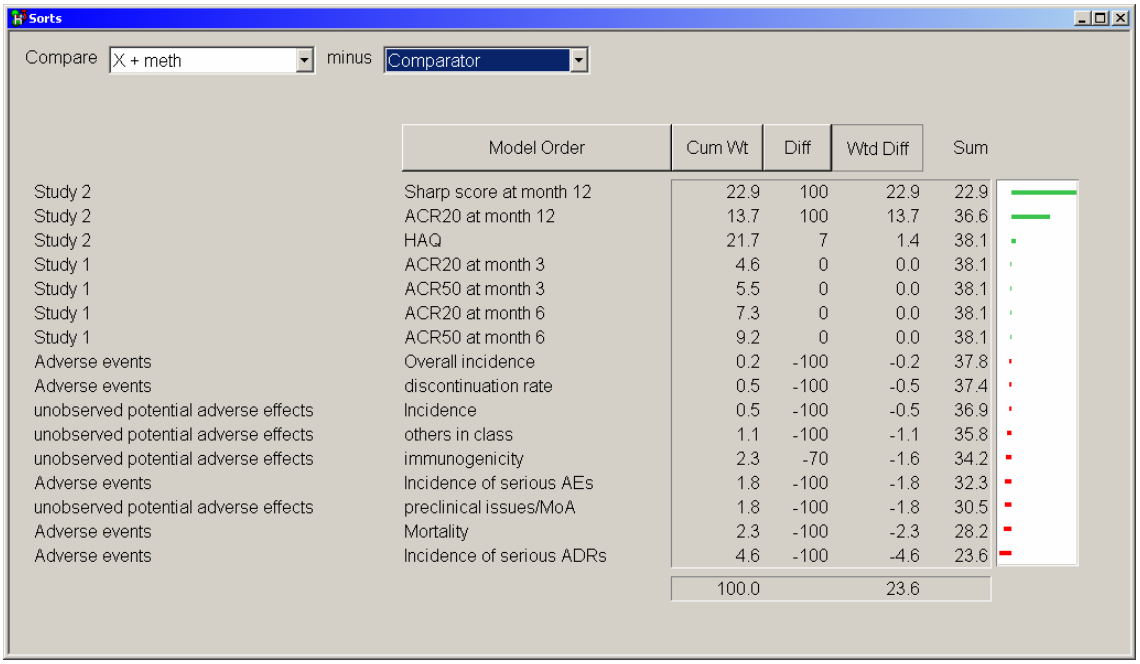


Figure 8: The criteria sorted in order of the biggest weighted difference in preference value between Product X plus methotrexate on the one hand and the placebo on the other.

The left two columns of words identify the 16 bottom-level criteria. The Cum Wt column shows the cumulative weight for each criterion. Each weight is the product of the normalised weights from the top of the value tree down to the criterion. This shows how a total weight of 100% is parcelled out amongst the 16 criteria. Each weight represents the difference in preference value between the least and most preferred options on the respective criterion; the weights do not reflect the absolute importance of the criteria. The Diff column shows the difference in the scores for the two options being compared for each criterion. For example, in Study 2, Product X plus methotrexate scores 100 on the criterion Sharp score at month 12, while the placebo scores 0, giving a difference of 100. Note that the difference in scores for those options on the HAQ is, however, only 7. The Wtd Diff column gives the weighted difference score, the product of Cum Wt and Diff. The Sum column gives the cumulative sum, with positive values associated with criteria favouring Product X plus methotrexate (green bars), and negative scores favouring the comparator (red bars). Note in the top matrix of Figure 3 that Product X plus methotrexate scores 47 while the comparator scores 24, for a difference of 23 (actually 23.6), the sum of the positive and negative differences in the Sum column.

The main take from this figure is that the key advantages of Product X plus methotrexate over the comparator are the Sharp score at month 12 and the ACR20 at month 12, two big differences on two highly weighted criteria. The biggest risk associated with the product is on the incidence of serious ADRs, but that risk, plus a number of smaller ones, still doesn't outweigh the benefits.

Conclusions

This modelling exercise, while somewhat hypothetical for the group, served to demonstrate several features of multi-criteria decision analysis as it might be applied to examining the benefit/risk balance for a medicine. First, it shows the importance of establishing the appropriate benefit and risk criteria. Secondly, it distinguishes between two types of judgement needed to establish the benefit and risk: assessments of the value of each option on the criteria, i.e. the process of scoring, from the relative weights of the criteria, i.e. the swing-weighting process. Third, it shows how a higher-level view can emerge from aggregating the benefits and the costs, a view that was not obvious at the level of the individual criteria. Fourth, it demonstrates the value of the social process, which engaged people with differing perspectives on the issues, using the MCDA model to serve as an agenda for a structured conversation and a tool for thinking. And finally, it demonstrated how the decision conferencing process can create aligned commitment in a group of key players who can now agree the way forward even if they don't agree about all the details.

Reference

Dodgson, J., Spackman, M., Pearman, A., & Phillips, L. (2000). *Multi-Criteria Analysis: A Manual*. London: Department of the Environment, Transport and the Regions. Download from www.odpm.gov.uk/stellent/groups/odpm_about/documents/page/odpm_about_608524.hcsp. Chapter 6 presents a brief overview of MCDA.

ANNEX 2

OVERVIEW OF THE CRITERIA AND CURRENT METHODS FOR BENEFIT-RISK ANALYSIS

Extracts from the presentation to the Workshop by Filip Mussen

Filip Mussen, Director, Regulatory Affairs Europe, Merck Research Laboratories, Belgium is also a Research Fellow at the Institute for Regulatory Science, working on a research programme to develop a model for benefit-risk assessment of new medicines. This has been part of a doctoral programme in association with the School of Pharmacy, Cardiff and Professor Larry Phillips from the London School of Economics.

Definitions in benefit-risk assessment

- *Benefit*: the proven therapeutic good of a product; should also include the patient's subjective assessment of its effects (WHO Collaborating Centre)
- *Risk*: the probability of harm being caused; the probability (chance, odds) of an occurrence (WHO Collaborating Centre)
- *Benefit* and risk are evaluative terms which contain value judgements (clinical studies cannot determine whether an effect is a benefit/risk, and how beneficial/harmful the effect is)
- *Benefit-risk balance*: more accurate than benefit-risk ratio (benefits and risks are not of the same nature)

Five concepts in benefit-risk assessment

- A separate benefit-risk balance for each indication
- All available data should be considered in benefit-risk assessment
- The nature of the disease should be taken into account for benefit-risk balance
- Absolute versus relative benefit-risk balance (compare with alternative therapies?)
- The benefit-risk balance is dynamic and evolves over time

Criteria to consider in benefit-risk assessment

The criteria selected are based on EU, FDA and ICH guidance. (The numbers in brackets refer to the 20 responses - 14 companies, 6 agencies - to questions in a CMR survey in 2002 that asked which factors should be included in a model for benefit risk assessment)

Benefit

For each pivotal trial:

- Efficacy (primary endpoint) versus comparator and its clinical relevance *(20/20)*
- Statistical significance of the efficacy results *(18/20)*
- Clinical relevance of the primary endpoints *(19/20)*
- Representativity of the studied population for the population targeted in the label *(18/20)*
- Evidence for the efficacy in relevant subgroups *(14/20)*
- Design, conduct and statistical adequacy of the trial *(18/20)*
- Confirmation of treatment effect by results of non-primary endpoints *(16/20)*

General benefit criteria:

- Confirmation of efficacy by results of relevant non-pivotal trials and extensions *(16/20)*
- Anticipated patient compliance *(11/20)*
- Clustering (consistency) of results of the pivotal trials

Risk

- Overall incidence of adverse effects (from clinical trials) *(16/20)*
- Overall incidence of serious adverse effects (from clinical trials) *(20/20)*
- Discontinuation rate due to adverse effects (from clinical trials) *(15/20)*
- Incidence, seriousness and duration of specific adverse effects (from clinical trials and post-marketing surveillance) *(20/20)*

- Interactions with other drugs and with food *(18/20)*
- Safety in subgroups (e.g., age, race, sex) *(20/20)*
- Potential for off-label use leading to safety hazards *(12/20)*
- Potential for non-demonstrated additional risk due to limitations of clinical trials and/or short market exposure
- Potential for non-demonstrated additional risk due to safety issues observed in preclinical safety studies but not in humans
- Potential for non-demonstrated additional risk due to safety issues observed with other medicines of the same pharmacological class

(the latter three criteria were previously clustered in the survey as 'generalizability of the safety profile to the general population' (18/20))

Why would models for benefit-risk assessment be useful?

- Enhance consistency in expressing the benefit-risk balance of a product
- Enhance objectivity in recommendations/ decisions on the benefit-risk of a product (by Registration Committees and in Marketing Authorization Applications)
- Increase transparency of regulatory decisions (approval and post-approval)
- Force the assessor to focus on benefits and risks
- Ideally, could be used as a tool to compare products
- Can be used as a tool for regulators and industry, but cannot substitute for the final decision-making

Objectives for a new model

A model which

- Is able to take into account the data in the MAA or otherwise available to regulatory agencies (i.e., safety and efficacy data from multiple clinical trials, post-approval AE data); No cost-benefit data
- Requires no additional analyses of source data (safety and efficacy), or meta-analyses
- Closely matches the current regulatory agency practices for benefit-risk assessment
- Can be used during initial registration and post-approval
- Can be validated
- Is applicable to all kind of drugs, including vaccines and OTC drugs

Which models are currently available?

Currently there are no well-established, validated models (qualitative or quantitative) although a few models are described in the literature:

- 'Principle of threes' (Edwards et al, 1996)
- TURBO model (Amery, 1998)
- Evidence-based benefit and risk concept (Beckmann, 1999)

These three models were mainly developed for pharmacovigilance purposes - post-marketing re-assessment (the 'Principles of Threes' model and the TURBO model are described in the CIOMS IV report)

Other models have been developed to assess the benefit-risk based on one clinical trial, e.g.,

- 'Benefit-Less-Risk Analysis' (Chuang-Stein)
- Mathematical model based on Numbers Needed to Treat (NNT) & Numbers Needed to Harm (NNH) (Schulzer & Mancini)
- 'Principle of threes' grading system (Edwards et al.)

Weaknesses of the current models

- Many criteria in the models are not well defined with regard to the type, quality and relative importance of the data to be taken into account
- Models do not take into account many of the benefit and risk criteria previously identified
- Models are not very sophisticated and allow only a very crude benefit-risk assessment
- Models have not been validated nor broadly used in practice

ANNEX 3

POINTS FROM THE WORKSHOP DISCUSSIONS

Historical perspective

Professor Stuart Walker provided an overview of the historical and present role of the Centre for Medicines Research in bringing together senior experts from the pharmaceutical industry and international regulatory agencies to address issues of policy and practice that impact on the development and evaluation of new medicines[1]. He noted that CMR had first addressed Risk Benefit assessment at a Workshop in 1985. Professor Walker emphasised the extent to which the subject has moved on by remarking that industry, at that time, was very resistant to economic issues being raised in relation to the assessment of medicines and somewhat dismissive of the concept that 'quality of life' could be a measurable factor in assessing the benefit of medicines.

In one important respect, however, benefit-risk assessment has not moved on. The re-evaluation of benefit-risk after authorisation is predominantly focused on risk – reports of adverse events and misuse – and the benefit of a medicine is rarely reviewed with a consequence that the benefit-risk almost inevitably appears worse with passing time.

Need for a benefit-risk model

The general discussions of the need for a viable benefit-risk model included the following points:

- The need for a better understanding of why different agencies come to different conclusions when faced with essentially the same application data;
- The increasing pressure on agencies to increase transparency and accountability and to establish a 'paper trail' to explain how decisions are reached;
- The need for a system that is sufficiently dynamic and flexible that it can be developed with experience with the potential that its application could be extended to include the views of a wider range of stakeholders, including patients and physicians;
- Acknowledgement that current approaches are somewhat haphazard not only on the part of the regulators, where decisions can be inconsistent but also on the part of companies where data in submissions on benefits and risks is not presented in a coherent and well structured manner.

Benefit goes beyond efficacy

It was agreed that discussions of the benefits of a medicine must encompass more than the measurement of efficacy in the clinical setting. Other factors include:

- Social settings and how the disease is managed in society, for example the added value of being able to treat a patient at home;
- The relative merits of a small improvement in a large number of individuals vs. a major improvement that is only seen in some individuals;
- The different perception of improvement that a patient and doctor may have. For example a relatively small clinical improvement in mobility may represent a major lifestyle change for the individual patient. Conversely clinicians might see great value in a drug that halts disease progress but this is of little benefit to the patient if the symptoms of the disease are not relieved.

There was also a discussion of the need to distinguish between the *efficacy* of a medicine, as evaluated in clinical trials and the *effectiveness* of a medicine that can only be judged once it has found its place in the clinical practice, post-authorisation.

Role of value judgements

It was acknowledged that value judgement is an integral part of benefit-risk assessment. Although it may be possible to draw up guidelines for the minimum data set that must be submitted to support a benefit-risk assessment, the decision making process will always include the application of informed opinion that is outside the scope of the data on its own. This is in contrast to the need to prove efficacy alone, which is often written into legal requirements. The MCDA model incorporates value judgement into the system of 'weighting' the values placed on different criteria used in the assessment.

Experts may disagree

History and experience has shown that there is no single 'correct' opinion on a matter as subjective as benefit-risk evaluation and there will always be disagreement among experts. Part of the purpose of introducing a model for assessing benefit-risk, however, is to move away from the scenario of discussions behind closed doors from which decisions emerge from 'smoke filled rooms' without explanations or accountability. A model that is applied transparently and consistently would also diminish the all-too-common scenario where a decision, apparently made by a committee of peers, is unduly influenced by vociferous and dominant individuals.

Perception of Risk

Assessment of risk cannot be based on scientific and statistical factors alone. The 'perception' of risk, especially by the public is extremely subjective and has been studied extensively. There was a discussion, at the workshop, of the concept that public perception of risk is based on three factors: Dread (e.g., BSE transmission to humans), Unknown (e.g., effects of MMR vaccination) and Numbers involved (e.g., motor accidents)[1]. Results of studies based on these factors show that, against all statistical arguments, the risks of human variant CJD are perceived as a much more serious than, for example, road accidents.

It was emphasised that training about making judgements on risk must include training on the nature of risk perception.

Models for efficacy

It was noted that, whilst there are regulatory guidelines on the data required to support claims for efficacy, there is no 'model' for assessing whether efficacy has been demonstrated. In particular, there appears little discussion on the number of patients that show improvement when treated vs the magnitude and nature of that improvement which could range from:

- Symptomatic relief;
- Slowing of disease progression; to
- Curing the disease

Role of comparators

Notwithstanding moves to harmonise the use of comparators in clinical trials, there remains a significant difference in practices between the US, where proof of efficacy against inactive placebo is accepted and Europe, where active comparators are expected in trials to prove efficacy and 'non-inferiority'.

There was a discussion of the issues involved when a new medicine is the first in its class and the implications for evaluating benefit and risk against active comparators with a significantly different mode of action. The question was also raised of whether judgements on the benefit of a first-in-class medicine are influenced by knowledge that similar drugs are in the pipeline.

Does the nature of disease impact judgements on benefits?

There was discussion on whether the seriousness of the disease is automatically a factor in assessing benefit. There is a tendency to regard drugs that benefit patients with a life-threatening disease, even when the improvement is only symptomatic, as having a greater value that a medicine for the treatment of a less severe or self-limiting condition. The example discussed was an anti-nauseant used for cancer patients undergoing chemotherapy and one recommended for sea-sickness.

Sequential decisions

Historically, as noted earlier, monitoring the way risk-benefit changes with time has focussed on risk and cumulative reports of adverse events and side effects. The benefit-risk assessment model should ensure that information on benefit is taken into consideration in a more balanced way when it becomes necessary to re-assess the safety of a product. An important factor is the increased trend towards 'conditional' approvals with commitments to carry out post-approval studies, requiring subsequent evaluation for any impact on the benefit-risk balance.

Effects of how regulators perceive their role

Not only do experts within one regulatory system disagree, but different conclusions are often reached between regulatory agencies. This can be, in part, the result of differing views of the role of regulation and the balance between responsibility to the individual patient and the population as a whole. Other factors may include:

- The impact and influence of the views of stakeholders including patients associations, physicians, health activists and reimbursement agencies;
- The application of the 'precautionary principle' resulting in very low risk tolerance;
- Political influences placed agencies when dealing with high-profile diseases (e.g., AIDS) and emotive issues where there is a degree of 'unknowns (e.g., MMR vaccine)

Communication of risk

Some medicines have known, serious adverse consequences for specific groups of patients (e.g., the elderly) or in pre-existing conditions. If the benefits, for eligible patients, are significant, can the 'risk' side of the equation be reduced by appropriate cautions and warnings in the product information (labeling)? There was general agreement that labeling and proposals for risk management and communication should be included in the model. It was also acknowledged, however, that experience has shown how difficult it is to ensure that physicians heed the warnings in product information.

Potential for off-label use

It follows, from the previous point, that the potential for off-label use must be a factor in deciding on the weighting to be assigned to specific and 'class' risks associated with a product. Off label use may not only relate to ignored warnings and contra-indications, however. It may take the form of using the product in extended, related indications, that are not explicitly included in the labeling. This may, in the long run, reveal new 'benefits' of the medicine.

Reference:

Slovic, P. (1987). Perception of risk. *Science, 236*, 280-285.