BENEFIT-RISK FRAMEWORK FOR THE ASSESSMENT OF MEDICINES:
Valuing the options and determining the relative importance (weighting) of benefit and risk parameters

13 DECEMBER 2012
PHILADELPHIA, USA

WORKSHOP REPORT
Workshop report authors
Stuart Walker, BSc, PhD, MFPM, FRSC, FIIBiol, FRCPath
Art Gertel, BSc, MSc
Neil McAslane, PhD
Lawrence Liberti, MSc, RPh, RAC
Patricia Connelly, BA, ELS

CIRS - The Centre for Innovation in Regulatory Science - is a neutral, independent UK-based subsidiary company, forming part of the Intellectual Property and Science business of Thomson Reuters. The mission of CIRS is to maintain a leadership role in identifying and applying scientific principles for the purpose of advancing regulatory and HTA policies and processes. CIRS provides an international forum for industry, regulators, HTA and other healthcare stakeholders to meet, debate and develop regulatory and reimbursement policy through the innovative application of regulatory science. It is governed and operated for the sole support of its members’ activities. The organisation has its own dedicated management and advisory boards and its funding is derived from membership dues, related activities and grants.

Centre for Innovation in Regulatory Science (CIRS)
The Johnson Building, 77 Hatton Garden, London, EC1N8JS, UK
Email: cirsci@cirsci.org
Website: www.cirsci.org

© 2013, Centre for Innovation in Regulatory Science, Ltd
Report date: May 2013
Section 1: Executive Summary

Background to the Workshop

At the annual CIRS Benefit-Risk Workshop in June 2012 there was an agreement from those who are developing benefit-risk methodologies for assessing medicines that there are four key stages to the evaluation: framing the decision; identifying the benefits and risks; assessing the benefits and risks; and interpretation and recommendation. Underpinning these stages was an overarching eight-step framework (Figure 1). This overarching approach has been termed by CIRS the Unified Methodologies for Benefit-Risk Assessment (UMBRA) Framework.

All the methodologies currently being developed by pharmaceutical companies and regulators incorporate these steps, whether explicitly or implicitly. This overarching framework provides the basis for common ground and agreement on the principles for benefit-risk assessment.

There are, however, two particularly challenging issues within the conduct of a benefit-risk assessment, one being the assessment of relative importance (weighting) and the other the valuation of the options such as medicines under investigation, comparators or placebo. With respect to determining relative importance, there is limited consensus with regard to the methods to be applied and a perception that the process can be highly complicated. Furthermore, whether and when relative importance should be assessed in a qualitative, semi-quantitative or quantitative process provides an area for open discussion. Valuing involves identifying the most appropriate measurements of a particular characteristic of a therapeutic option and developing methods that permit consistent comparisons across these values.

This meeting brought experts from the pharmaceutical industry together to debate and discuss the two critical issues faced when using a benefit-risk framework: the weighting (establishing the relative importance) and valuing of the options. These were discussed from the perspective of their utilisation by both medicine developers and regulatory authorities.

Workshop Objectives

- Provide a forum for industry to review methods and discuss the utilisation of weighting and values in making a benefit-risk decision
- Assemble industry perspectives as to which weighting and valuing standards could be used by regulatory agencies and which issues need to be addressed within the benefit risk-framework
- Present the opportunity for attendees to put forward proposals for discussion at the CIRS annual benefit-risk Workshop, scheduled for June 2013, on the principles that should be applied in these steps of the framework

Key points from presentations

SESSION: RELATIVE IMPORTANCE FOR BENEFIT-RISK ASSESSMENT — WHAT ARE THE METHODOLOGIES?

From its initiation, the goal of the CIRS benefit-risk programme has been the development of an internationally acceptable, structured, systematic standardised approach to the benefit-risk assessment of medicines. CIRS Founder, Professor Stuart Walker set the scene for the day’s discussions by citing the ongoing work of CIRS in this field including the facilitation of the Consortium for Benefit-Risk Assessment (COBRA) in their efforts to develop a common benefit-risk assessment template to enable shared
regulatory review as well as the development of the overarching eight-step UMBRA framework, initiated during a discussion at the CIRS annual Benefit-Risk Workshop in June 2012.

The number, variety and complexity of methods used for weighting the attributes of new medicines may present barriers to their use. In addition, there are no standards or guidance as of yet for the use of the methodologies, which can be regarded by regulators as novel and subject to the introduction of industry bias.

Dr Bennett Levitan, Director, Epidemiology, Janssen Research & Development, USA presented information regarding selected methods for weighting, summarising the advantages and disadvantages of the models in terms of their theoretical justification, the time and expertise needed to implement and the ease with which the results are communicated.

Although formal quantitative methodologies, such as multi-criteria decision analysis can be used in certain scenarios for benefit-risk assessment, valid conclusions can also be drawn based on qualitative benefit-risk assessments.

Dr Elias Kouchakji, Executive Medical Director and Dr Qi Jiang, Executive Director, Global Biostatistical Science, Amgen, USA discussed a hypothetical case study of the use of several qualitative approaches to the benefit-risk evaluation of an angiotensin-converting enzyme inhibitor and a control treatment of congestive heart failure. An overall positive benefit-risk profile was demonstrated for the ACE inhibitor, based on multiple considerations including the magnitude of treatment effect and the weights and utility assigned to the benefit-risk endpoints.

In swing weighting the benefit-risk parameters of a new medicine, decision makers assign the highest “score” to a change in the potential range of measurement of the parameter that they judge to be of greatest importance and then assign scores to the potential changes in the other parameters relative to their importance in comparison with that most important change.

Dr James Felli, Research Fellow, Eli Lilly, USA, explained that a substantial portion of the value of evaluation methodologies such as swing weighting lies in the opportunity they afford to encourage critical conversations among stakeholders as to such issues as the differences in individual perspective, the choice of comparators and the appropriateness of scale in decision making.

Natalizumab was approved in 2004 by the FDA for the treatment of relapsing remitting multiple sclerosis. Although the drug was removed from the market because of an associated incidence of progressive multifocal leucoencephalopathy, it was later reintroduced because of patient demand. Dr Diana Hughes, Vice President, Worldwide Safety Strategy, Primary Care Business Unit Lead, Pfizer Inc, USA reviewed an evaluation of the benefits and risks for natalizumab using the six-step approach of the Pharmaceutical Research and Manufacturers of America Benefit-Risk Assessment Team (PhRMA BRAT). Because the analysis also included the elicitation of weights through linear additive swing weighting, it was possible to use graphic displays such as a waterfall chart to illustrate the individual contribution of the outcomes to the ultimately positive benefit-risk profile for natalizumab.

The benefit-risk evaluation group at GSK works to embed their approach for the measurement and articulation of the key benefits and risks of new medicines within product teams for their use at global safety board milestone reviews of new medicines and beyond.

Dr Marilyn Metcalf, Director, Benefit Risk Analysis USA GlaxoSmithKline, USA discussed a fictionalised case study that illustrated the team’s process of using a value tree to facilitate the framing of the medicine’s evaluation in terms of the disease context, the methods for measuring benefits and risks, the differences in sub-group effects and the reversibility and monitorability of adverse events and graphs that display the drug’s benefits and risks side by side and that serve as a backdrop for informed discussion.

Dr Carmen Bozic, Senior Vice President and Global Head, Safety and Benefit-Risk Management, Biogen Idec presented the results of the use of Markov modelling and conjoint analysis in the assessment of the benefits and risks of natalizumab in the treatment of relapsing and remitting multiple sclerosis. Conjoint analysis results, which were included as part of a variation submission to EU regulators, revealed that in return for a natalizumab-level efficacy, patients were willing to accept a 3.8 per 1,000 annual risk of death or severe disability due to progressive multifocal leucoencephalopathy or 7.6 per 1,000 risk over 2 years or 15.2 per 1,000 risk over 4 years.

Because multiple initiatives have tested the operational and scientific validity of frameworks and quantitative methodologies and methodological approaches to weighting, such as those used in multi-criteria decision analysis are relatively well understood, Dr Becky Noel, Senior Research Scientist, Eli Lilly, USA believes that
it is reasonable to assume that methodology is not the only or even the primary impediment to weighting and valuing in decision making. She explained that recognition is required that the contentiousness surrounding weighting cannot be addressed by focusing on the numbers alone. In addition to any methodological work that may be needed to advance the science of weighting, the policy and operational requirements for implementing a weighting process within an organisation must also be developed.
### Session 1: Benefit-risk assessments: relative importance for benefit-risk assessment — what are the methodologies?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chairman’s introduction</td>
<td>Prof Stuart Walker, Founder, CIRS</td>
</tr>
<tr>
<td>Principles, rationale and types of methodologies for assessing relative importance</td>
<td>Dr Bennett Levitan, Director, Epidemiology, Janssen Research &amp; Development, USA</td>
</tr>
<tr>
<td>Case studies of different methodologies for relative importance:</td>
<td></td>
</tr>
<tr>
<td>A ranking/qualitative approach</td>
<td>Dr Elias Kouchiaki, Executive Medical Director and Dr Qi Jiang, Executive Director, Global Biostatistical Science, Amgen, USA</td>
</tr>
<tr>
<td>The swing weighting approach</td>
<td>Dr James Felli, Research Fellow, Eli Lilly, USA</td>
</tr>
<tr>
<td>The IMI Tysabri evaluation</td>
<td>Dr Diana Hughes, Vice President, Worldwide Safety Strategy, Primary Care Business Unit Lead Pfizer Inc, USA</td>
</tr>
<tr>
<td>Roundtable discussion and development of recommendations to regulators: What are the potential methodologies for regulatory assessment?</td>
<td></td>
</tr>
</tbody>
</table>

### Session 2: Benefit-risk assessments: valuing the options for benefit-risk assessment — what are the approaches?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valuing the options</td>
<td>Prof Stuart Walker, Founder, CIRS</td>
</tr>
<tr>
<td>Focus on valuing the options in a benefit-risk framework</td>
<td></td>
</tr>
<tr>
<td>A company case study</td>
<td>Dr Marilyn Metcalf, Director, Benefit Risk Analysis USA, GlaxoSmithKline, USA</td>
</tr>
<tr>
<td>A company case study</td>
<td>Dr Carmen Bozic, Senior Vice President and Global Head, Drug Safety and Benefit-Risk Management, Biogen Idec, USA</td>
</tr>
<tr>
<td>Roundtable discussion and development of recommendations to regulators: What are the potential methodologies for regulatory assessment?</td>
<td></td>
</tr>
<tr>
<td>Should relative importance and valuing be combined in making the benefit-risk assessment for company submission or agency review?</td>
<td>Dr Becky Noel, Senior Research Scientist, Eli Lilly, USA</td>
</tr>
</tbody>
</table>
Section 2: Roundtable Discussions

Roundtable 1: What are the potential methodologies for regulatory assessment? Discussion and recommendations

Group A
Experience with benefit-risk assessment was highly variable within the companies represented in Group A. Some companies have included a detailed benefit-risk analysis in recent regulatory applications. Other companies are developing internal guidelines to help select the most relevant method for benefit-risk analysis of a particular medicine. Participants agreed that in all cases organisations needed to consider the real-world practicality of using a selected method within their group because not every benefit-risk decision requires a complicated or sophisticated methodology. Despite the relative informality of most of these analyses, however, the assessments were noted by regulatory authorities either in their own review or as material for an Advisory Committee. Other companies are currently not performing any formal quantitative benefit-risk analysis.

Recommendations
1. Any method for regulatory assessment of benefit-risk would have to have theoretical support to be acceptable to regulatory authorities, ruling out informal methods such as direct entry. Health authorities, who are already implicitly performing zero-or-one assessments, should at a minimum explicitly perform and communicate their benefit-risk categorisations.
2. The same overarching framework for benefit-risk assessment should be used at agencies across the world, even though approaches to encourage the use of these assessments within agencies may be variable because of the varying level of progress in and acceptance of benefit-risk evaluation.
3. The US FDA should use the upcoming patient workshops as an opportunity to gather information about patient benefit-risk preferences.
4. While conjoint analysis can be helpful, rather than use this for every product it could be used on a case-by-case basis based on an agency request.

Group B
Recommendations
1. Regulators should consider using a weighting approach, which will help to document an explicit, transparent decision-making process and correct any potential overlap of particular attributes in the model. In some cases two different weighting methods could be used to test the robust nature of decisions.
2. Regulators should also consider using a toolbox approach, as one method will not fit all circumstances and all products. With this approach, some straightforward evaluations may not require a weighting method at all whereas a simple method such as ranking could be used for less complex evaluations and the methods could progress in complexity as required.
3. Patient preference data should be built into clinical trial endpoints as well as in the weighting of those endpoints in product benefit-risk assessment during regulatory review.

Group C.
Recommendations
1. Although informal and implicit weighting is already performed by all organisations, transparency in this weighting is required to communicate an understanding regarding what benefits and risks are considered by each stakeholder to be important and why.
2. Weighting should be used to inform but not dictate decisions and should support but not override clinical judgement.
3. This Group also recommended a toolbox approach in which different methods are used situationally.
   a. They also recommended that those methods be standardised and consistent and that the preferred method for weighting be determined through discussion between agency and sponsor as a product is developed.
   b. Among weighting methods in the toolbox, the Group recommended that swing weighting and direct assessment be included. Although direct assessment does not have an axiomatic foundation, it does
have a behavioural justification.

4. Because weighting is subjective, considerations should include both patient-reported and clinical outcomes and also include patient preferences in a meaningful and substantive manner.

Additional commentary

- To eliminate regulatory objections to patient-reported outcomes developed by sponsors, preference surveys should not be designed by companies to test a particular drug, but be developed by impartial parties with proper academic expertise for use across a therapeutic area.

- The Critical Path Institute has formed a consortium of industry members working in cooperation with the US FDA, who are developing patient-reported outcome instruments for a number of disease states. More information about the Consortium can be found here.

- The FDA benefit-risk grid does not include the potential for ranking or weighting attributes. It would be helpful to know what methods are being used for these evaluations, that is, what questions are being asked when evaluators fill in the grid.

- Although PhRMA and others proposed that the FDA select simple disease states with clearly delineated treatment endpoints for the upcoming patient workshops, it appears that the FDA selected 20 complex disease areas, with the idea that the processes that are developed from these workshops would be extendable to a larger set of more readily defined conditions.

- Industry teams that engage in decision analysis exercises enhanced a team's ability to develop clear, succinct opinions regardless of the outcome. Even though quantitative results are not always possible, the use of good decision analysis methods augments qualitative thinking; however, not all teams are amenable to this type of discussion and debate.

- Although it would seem logical for sponsors to develop benefit-risk assessments for a drug in specific subgroups for which the profile would be most favourable, it may not be a viable investment. For example, it costs approximately $400,000.00 to conduct a 300-patient preference study. It would be challenging to partition that population into two subgroups from whom statistically significant results could be derived and further refinements would result in point estimate losses in data. The only way that it may be possible to develop such a conjoint analysis would be through extensive voluntary participation.

- Regulators may narrow a patient population on a selective basis as part of an approval process, even though the data from these smaller subgroups may not have been considered adequate for submission by the sponsors. Regulators may also ask sponsors for a subgroup analysis if the benefit-risk profile in a large population is not definitive.

- Health technology assessors also narrow patient populations when they determine for which portion of a total population a treatment is reimbursable. How they use a benefit-risk approach to this decision requires investigation.

Roundtable 2: Quantitative, qualitative or hybrid?

Group A

Group A spent time developing definitions to ensure common points of reference in their discussion. Ultimately, the definition of value that was used is the numerical representation of a factor that can support a preference and weighting was defined as the importance of a factor in a benefit-risk assessment.

It was the consensus of the Group that a hybrid qualitative/quantitative approach to benefit-risk evaluation is required that identifies, values and weights certain key attributes through discussion. Most companies are not yet in a position to use this approach, but do include elements of it in their deliberations. Results of these deliberations are used to develop the Overview section of the dossier.

Group B

This Group identified reluctance on the part of regulatory agencies to accept quantitative assessments, although the agencies do seem to enthusiastically accept the visualisation aspect resulting from quantitation. For their part, companies seem to primarily use quantitation internally, especially in preparation for Advisory Committees, possibly because they are seen as critical labour-intensive events which can benefit from creative approaches to illustrating
scientific data. On the other hand assembling a file for a traditional submission is regarded more as an occasion to use a narrative approach, saving visualisations for challenges issued by the regulatory agency.

The Group also observed a trend toward gradual acceptance of weighting being achieved in a stepwise fashion within some companies, as training exercises that incorporate quantitative assessments are preparing teams to begin to use these approaches. These exercises may be accompanied by internal guidances or white papers on the topic.

Group B agreed that a hybrid approach to benefit-risk assessment is needed and posited that companies will likely select the methods that best suit their needs and that the selection will likely vary on a product-by-product basis.

In a general group discussion it was noted that companies would be more amenable to invest the time and effort required to use the simple software to create benefit-risk visualisations that is being developed as part of UMBRA BRASS-2 (which incorporates value tree and visualisation elements of the PhRMA BRAT approach) if regulatory agencies first indicated a preference for benefit-risk visualisation as part of submissions. This may happen gradually, but visualisations presented at Advisory Committees are now being used by the regulatory agencies themselves in their summary documents and briefing books.

**Group C**

This Group also spent time defining weighting and valuing. They agreed that most companies, while not formally using the terminology **valuing**, are evaluating the clinical significance of events and how they relate to available therapies through the use of specific measurements. Whilst no quantitative measurement for weighting each event is described formally in dossiers, it was agreed that all qualitative discussions are based on some basic form of quantitation. The Group concurred that a hybrid methodology is most likely to meet various stakeholder needs simultaneously.
Benefit-risk framework for the assessment of medicines: Valuing the options and determining the relative importance of benefit and risk parameters

Professor Stuart Walker
Founder Centre for Innovation in Regulatory Science

From its initiation, the goal of the CIRS benefit-risk programme has been the development of an internationally acceptable, structured, systematic, standardised approach to the benefit-risk assessment of medicines. Toward that end CIRS has facilitated the work of the Consortium for Benefit-Risk Assessment (COBRA), a group of regulators from Health Canada, Singapore, Australia and Switzerland who are developing a common assessment benefit-risk template to enable shared regulatory reviews. Using the EMA Benefit-Risk Guidance Document of 2008 as its basis the group produced an electronic template for the documentation of benefit-risk assessment that uses a qualitative or semi-quantitative approach (Figure 2). A 2010 feasibility study and a retrospective pilot study in 2011 demonstrated this approach and the results of a recent prospective study will be presented at the June 2013 Benefit-Risk Workshop in Washington DC.

Other groups, notably the US FDA, EMA and the Pharmaceutical Research and Manufacturers of America Benefit-Risk Assessment Team (PhRMA BRAT) have developed benefit-risk evaluation methodologies, all of which have certain elements in common. At the CIRS June 2012 Annual Benefit-Risk Workshop, a Syndicate discussion group proposed an overarching eight-step framework under which all of these methodologies could be organised (Figure 3). This framework was subsequently endorsed by the CIRS Benefit-Risk Taskforce. UMBRA – Unified Methodologies for Benefit-Risk Assessment – is a platform established by CIRS in 2012 for the coordinated development of benefit-risk assessment. The Taskforce, which comprises eight regulator and eight industry members was designed to facilitate the exchange of information among stakeholders.

Through UMBRA, CIRS is committed to the continued exploration of the topic of benefit-risk. Interest from numerous agencies has resulted in CIRS forming the Southeast Asian Benefit-Risk Evaluation (SABRE) group, which will begin piloting the first version of the Benefit-Risk Assessment Support System (BRASS-1), which is the combination of the COBRA electronic pro forma template and its User Manual and Glossary. Following discussion with the EMA Benefit-Risk Taskforce, CIRS will also make BRASS-1 available to interested European agencies and companies for piloting.

Finally, Professor Walker noted that CIRS will hold four Workshops on the topic of benefit-risk during 2013.

13-14 March: The patient voice in clinical development: Can patients contribute to the benefit-risk assessment of new medicines? Surrey, UK

20-21 June: Implementing an internationally acceptable framework for the benefit-risk assessment of medicines: How close are we to this objective? Washington, DC, USA

2-3 October: Is there a commonality across the structured decision frameworks used by HTA and regulatory agencies? UK

Week of 9 DEC: Technical Forum: Exploring ways to maximise the value of Periodic Benefit-Risk Evaluation Reports (PBRERs): Addressing agency needs and company approaches to post-approval benefit-risk assessment Philadelphia, USA
Professor Walker set the scene for the day’s discussion by citing two of the participants of the CIRS annual Benefit-Risk Workshop in June 2012 who summarised the requirements and potential value of a framework to evaluate the benefits and risks of new medicines.

“...a framework that accurately and concisely describes the benefit and risk considerations associated with medicines will help [FDA] reviewers apply a structured approach in regulatory decision making.”

Dr Theresa Mullen, Associate Director, Office of Planning and Informatics, CDER

Dr Jason Ferla, Director, Prescription Medicines Clinical Unit 3, Therapeutic Goods Administration

Reference


Interest from numerous agencies has resulted in CIRS forming the Southeast Asian Benefit Risk Evaluation group, which will begin piloting the first version of the UMBRA Benefit-Risk Assessment System (BRASS-1)
**Principles, rationale and types of methodologies for assessing relative importance**

Dr Bennett Levitan  
*Director, Epidemiology, Janssen Research & Development, USA*

**Why is there resistance to the formal weighting of benefits and risks?**

Medical decision making is typically performed through the intellectual integration of data on the part of the decision maker and is communicated verbally rather than by the use of a numeric system such as may be a part of a formal weighting methodology. In addition, value judgements, such as those used in the weighting of benefit and risk parameters, may be regarded as “less scientific.” Moreover, some stakeholders may hold the view that a quantitative approach to benefit-risk weighting provides an answer rather than serving as a means to obtain clarity. Finally, the number, variety and complexity of methods used for weighting may present further barriers to their use and there are no standards or guidance as of yet for the use of the methodologies, which can still be regarded by regulators as novel and subject to the introduction of industry bias.

**Common methods to obtaining weights**

There are five broad categories of methods to obtain benefit-risk weighting: informal, tradeoffs and allocations, stated choice methods, pairwise methods and the broad catch-all “others.” Dr Levitan provided background on selected methods from three of those categories.

**Informal methods**

- **Zero/One weighting:** The identification of relevant outcomes is one of the first steps in which the values of the decision maker come into play, as outcomes are weighted as relevant or not. This assessment is performed implicitly when developing a clinical protocol, statistical analysis plan, value tree or benefit-risk approach.

- **Categorisation:** Decision makers or clinical experts assign each endpoint to a category in an n-point scale. Existing, validated scales such as the Common Terminology Criteria for Adverse Events can be used.

**Direct entry:** Decision makers or clinical experts review endpoint or attribute definitions and generate a numeric weight. They may use a heuristic similar to tradeoff or allocation approach, but this method is inherently unstructured and consequently may be difficult to explain and defend.

**Tradeoff and allocation methods**

**Point allocation:**

1. Decision makers start with well-defined attributes; for example, \( \text{headache relief} = \) reduction from severe or moderate pain to mild or no pain in two hours; \( \text{rapid onset} = \) reduction from severe or moderate pain to mild or no pain in one hour and \( \text{myocardial infarction (MI)} = \) the number of MIs per 1,000 patient-years.

2. Next, potential incremental changes in the attributes are defined; for example, 1% increase in headache relief; 1% increase in rapid onset; an increase in 1 MI per 1,000 patient-years.

3. The incremental change that has the greatest impact on decision making is then selected and assigned 100 points.

4. Values between 0 and 100 are allocated to the incremental changes in the other attributes, reflecting their clinical importance relative to the attribute with the greatest impact.

5. Finally all weights are scaled back so that they equal 100.

6. The overall results can then be visually portrayed to stakeholders to ensure buy-in.

This approach is similar to the ‘100 coin’ approach in which incremental changes in attributes are assigned a portion of 100 coins according to their clinical value.

**Swing weighting:** This approach is similar to point allocation, but is based on a full range of attributes. It is critical to specify the range for each attribute that is relevant to the decision; for example, the proportion patients with headache relief (reduction from severe or moderate pain to mild or no pain within 2 hours after treatment) should range from 20% to 80% and the number of MIs per 1,000 patient-years (MI defined per clinical criteria) range from 0 to 40.
Decision makers identify the attribute whose “swing” from lowest to highest value in its range would have the greatest impact on the decision, which in this case would be the number of MIs per 1,000 patient-years from 0 to 40. For each other attribute, an assessment would be made of the fraction of this value that would be achieved by swinging the other attribute from its lowest to highest value; for example, the proportion patients with headache relief swinging from 20% to 80% would have 1/5 the impact on the decision as the swing for MI.

**Stated choice methods**

**Background for stated choice methods**

Stated Preferences are those preferences elicited by using hypothetical situations; for example, asking a decision maker whether Treatment A or B is preferable, while Revealed Preferences are those preferences that are revealed by choices in real-world situations such asking a decision maker whether Treatment A or B was selected. Although the use of Revealed Preference data might appear to be most advantageous, Stated Preference surveys can assess treatments that are not yet available whilst Revealed Preference data suffer from many confounders such as the effects of differing insurance plans and access and decisions made by other stakeholders. Furthermore, there may be little variability in key treatment attributes being surveyed and these data may be difficult and time consuming to obtain.

**Conjoint analysis surveys:** These surveys are designed with clinical experts and tested repeatedly before use. Typically, several hundred subjects are asked to make a series of choices between hypothetical treatments that are based on different combinations of attributes; for example, a treatment that is associated with an 8% chance for a disabling stroke or 12% chance for bleeding requiring transfusion. Surveys include training on attribute definitions and treatments and the available choices are reminiscent of real-world medical decisions. Surveyors may also vary the combination of attributes included and note the different choices made by respondents. Choices made by survey respondent are converted to weights by a regression model.

Data obtained from a triptans preference study conducted by Pharmaceutical Research and Manufacturers of America Benefit-Risk Assessment Team/Next Steps Working Group (PhRMA BRAT/NSWG) revealed that the most important attribute associated with two different migraine treatments was the risk of MIs. Excluding that risk, other potential adverse events for the two treatments were regarded as roughly equivalent and three times less important than the potential benefit of preventing limitations on everyday activities. The survey results suggested that the maximum acceptable risk that patients would be willing to accept to relieve activity limitations during migraines was a 2/1000 annual chance of MI.

The mathematical formula underpinning conjoint analysis methods and similar approaches is that the utility of a treatment is the sum of the utilities for the components. For example, the utilities for treatment A, which results in no pain at 2 hours and 20% chance of relapse at 24 hours and so on … and treatment B, which results in no pain at 2 hours and 0% chance of relapse at 24 hours and so on … would be represented as:

\[
U_A = U_{\text{no pain at 2 hours}} + U_{\text{20\% chance relapse}} + \ldots \\
U_B = U_{\text{no pain at 2 hours}} + U_{\text{0\% chance relapse}} + \ldots 
\]

In this case, the probability of picking treatment A is proportional to the exponent of the utility A over the sum of the exponent of the other utilities.

\[
p(A) = \frac{e^{U_A}}{e^{U_A} + e^{U_B}}
\]

A regression model identifies the component utilities that give the choice probabilities that best match the choices that were made, which with proper rescaling become the preferences elicited by the survey.

**Best-Worst scaling:** This method is an increasingly popular means of stated preference elicitation in health applications. It is similar to conjoint experiments, although it can be easier for subjects to use, especially in evaluating a treatment with many attributes. The analysis and results are similar to those for conjoint analysis.

**Common methods to apply weight**

Although the number of endpoints and their variety of clinical significance may present a challenge to the use of value trees in the evaluation of benefits and risks, a focus on events of comparable clinical impact lessens the need for formal weighted methods.

A forest plot, a visual representation of weight ranking for a study drug and its comparator...
is a compromise between quantitative and qualitative methodologies. Once weightings have been established for outcomes for the two therapies, a forest plot can be developed that organises the outcomes vertically by their weighted importance and along a horizontal scale of risk of occurrence per common basis (ie, per 10,000 patients), with results on one side favouring the study drug and results on the other side favouring the comparator (Figure 4).

**Maximum acceptable risk** uses a ratio of weights to give a threshold for acceptable tradeoffs; for example, the previously cited survey that suggests that patients will accept a 2/1000 annual chance of MI to relieve activity limitations during migraines. It is based on the similar mathematics for additive linear models as in multi-criteria decision analysis (MCDA) or incremental net clinical outcome models, but uses clinical language to express the result.

**Incremental and net clinical outcome** represents the difference in the weights of outcomes times the rate of their occurrence between two treatments. There are many ways to visually represent the use of this methodology, including showing the contribution of individual components, but essentially the outcome is a single number that needs to be translated into a utility.

In the **stochastic MCDA** model used by Innovative Medicines Initiative Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (IMI PROTECT) for an analysis of the multiple sclerosis drug natalizumab, a visual representation was developed to demonstrate the results of weighting and valuing endpoints of two therapies. This model is similar to that used to develop forest plots but employs random, non-binary endpoints and accounts for uncertainty.

Certain limitations are associated with the methodologies. Although offering good approximations, the total values achieved in additive models can be incorrect. The impact of an event may not be proportional to its frequency or severity as is assumed in some of these approaches. Using weights for a health state with a duration that is different than that used in the weight assessment can give misleading results. The advantages and disadvantages of the models in terms of their theoretical justification, the identity of the parties assessed, the time and expertise needed to implement and the ease with which the results are communicated (Figure 5) all affect their utility according to various stakeholder needs.
Benefit-risk assessments: A ranking/qualitative approach

Dr Elias Kouchakji  
Executive Medical Director

Dr Qi Jiang  
Executive Director, Global Biostatistical Science, Amgen, USA

Dr Consuelo Blosch, Dr Laura Bloss, Dr Chunlei Ke

Evaluation of new treatments has always involved a benefit-risk (B-R) assessment. These evaluations have historically been informal and often subjective. Companies, regulatory agencies and other governance bodies are increasingly implementing structured B-R assessments. There are difficulties inherent in B-R assessment, however, including multiple, varied perspectives regarding the relative importance of a medicine’s benefits and harms and the changing nature of the assessment over time.

The distinctions between qualitative and quantitative approaches for B-R assessment are not very clear. Quantitative approaches require a qualitative framework for developing models. Qualitative approaches may be sufficient for simpler B-R decisions. Qualitative B-R assessment methods refer to a structured framework such as the CIRS framework, the BRAT framework, PrOACT-URL, or the FDA framework.

Ranking or weighting the individual benefits and risks is the key to B-R assessment for both qualitative and quantitative approaches.

In the CIRS Unified Methodology for Benefit-Risk Assessment (UMBRA) framework, the eight steps of the evaluation are broken into four key stages: framing the decision, identifying the benefits and risks; assessing the benefits and risks and interpretation and recommendation. Among these eight steps, refining the value tree and assessing relative importance of benefits and risks are the two key steps in B-R assessment.

Weighted approaches account for the nature and importance of different benefits and risks. They facilitate the use of subjective weighting and objective value data and allow for the viewpoints of multiple stakeholders. Some methods are less formal while others are more formal. For example, less formal weighting approaches include ranked outcomes, in which parameters are assigned predefined categories of importance; point allocation and equal weighting. Formal approaches comprise methods such as utility weighting; patient-preference weighting (e.g., conjoint analysis), and multi-criteria decision analysis (MCDA).

Currently there are no commonly accepted regulatory standards or guidance on how to select weights. The CIRS framework provides the prerequisites for applying whichever approach serves the stakeholders’ scientific needs without specifying a particular method for weighting. Amgen is currently evaluating both qualitative and quantitative approaches. Amgen was invited to discuss their thoughts/practice on Ranking/Qualitative approaches during this Workshop. Therefore, the focus of this presentation was evaluating a B-R Assessment using a Ranking/Qualitative approach.

Case study

Background

In patients with congestive heart failure (CHF), the heart is unable to adequately pump blood and patients may experience dyspnoea, fatigue or oedema. CHF is associated with reduced life expectancy and represents an unmet medical need. In this hypothetical case study, non-real data are presented for two treatment groups, patients receiving an angiotensin-converting enzyme (ACE) inhibitor and those receiving a placebo.
Potential benefits of CHF treatment include a delay in occurrence of hospitalisation for heart failure, an improvement in symptoms as measured by New York Heart Association (NYHA) class, and a prolongation of overall survival (OS). Risks, meanwhile, include angioedema, cough, renal toxicity and diarrhoea. These benefits and risks can be represented in a value tree, and once the hazard ratios are calculated, displayed as a forest plot, with data on the left favouring treatment with the ACE inhibitor and data on the right the placebo (Figure 6).

In a graphic display of point allocation in this case study, a table was created that listed the benefit and risk parameters associated with the ACE inhibitor and placebo treatment groups in rank order. Point allocations assigned to these parameters indicated not only the relative importance of the parameter but also the magnitude of importance and were used to develop adjusted risk difference per 100 patient years. (Figure 7) Point allocation is a simple and commonly used approach in which a number of points are applied to each parameter based upon reasonable medical and scientific judgement. It may be argued, however, that this method introduces subjectivity to decision making and is not scientifically well grounded.

Displayed in a forest plot, these adjusted risk differences demonstrate a positive benefit-risk ratio for the ACE inhibitor (Figure 8).

In another method of assessment, the numbers needed to treat (NNT) to achieve benefit and the numbers needed to experience harm (NNH) can be calculated. However, these calculations presume that all benefit and risk parameters for the treatments carry an equal weight and can be adjusted for utility (Figure 9).

Finally, an effects table was created that presents the hypothetical study results for the endpoints as favourable and unfavourable effects, with units of measurement and a possible range of values. Although this table shows an absolute risk increase for the occurrence of angioedema with ACE inhibitor treatment as with the other methods of assessment, the results can be seen to favour ACE inhibitor treatment (Figure 10).

**Discussion and Conclusions**

B-R assessment for the hypothetical ACE inhibitor and placebo treatment groups showed that the effects of each treatment on survival were not different, although a significant delay in time to first hospitalisation for heart failure was demonstrated with the trial drug. Because of the low importance assigned to cough and diarrhoea, these unfavourable effects did not significantly impact the B-R assessment. There was some increase of angioedema with the study drug and this was considered the key risk of treatment. Overall, a positive B-R profile...
The selection of endpoints for B-R assessments and their correlation with one another is critical. In fact, the pre-selection of those endpoints, while challenging to accomplish, is optimal. It should be recognised that the endpoints are likely to vary among different disease areas and both their clinical and statistical significance must be considered. There are challenges associated with the determination of the relative importance of B-R parameters through reasonable medical judgement. Because understanding of the process and results of weighting is variable among stakeholders and consensus is limited, transparency is essential. It must be remembered that all outcomes are not of equal significance and patient perspectives of those outcomes may differ from other decision makers. Sensitivity analyses will continue to be required to facilitate further understanding of the B-R assessment.

Although formal quantitative methodologies, such as multi-criteria decision analysis can be used in certain scenarios, the qualitative approach can be strengthened by the use of quantitative methodology in many assessments and valid conclusions can be drawn based on qualitative B-R assessments, which are essentially qualitative assessments of quantitative data.
The swing-weighting approach

Dr James Felli
Research Fellow, Eli Lilly, USA

Weighting is the determination of how much significance each of several factors or attributes will contribute to one’s assessment of the value of an alternative. For example, in Type 2 diabetes, there are seven attributes associated with treatment alternatives: reductions in the levels of post-prandial glucose, haemoglobin A1C, low-density lipoprotein, hypoglycaemia, oedema, increases in the levels of high-density lipoprotein and vitality.

After providing a potential range of scores for changes in these attributes, decision-makers can determine what “swing” over the potential range of scores is of greatest importance and assign an arbitrary score to this swing. For example, if a decision-maker considered a swing in post-prandial glucose within the range of 130-170 mg/dL to be the most important change in attributes among all the potential changes associated with a treatment alternative, this change could be assigned a score of 100. Next, scores are assigned to the potential changes in other attributes relative to their importance in comparison to the change that was scored as 100.

It should be understood, that an attribute that is scored as zero may be intrinsically important, but cannot be established as a point of differentiation in the treatment being evaluated and thus has no operational value in the decision-making process. For example, in this case, the attribute of vitality was scored as zero because all diabetes therapies affect vitality in the same way. Similarly, although a reduction in haemoglobin A1C may be regarded as an intrinsically important attribute of diabetes therapies, if most therapies affect this attribute in a similar way, its differentiation and consequently, its operational importance as an attribute is not as high as might be expected (Figure 11).

All of the scores are then totalled and the weights are normalised relative to that number to [0,1]. These developed weights reflect the intrinsic and operational value of the attributes according to the evaluator, but these values must still be validated through stakeholder review (Figure 12).

Methods for and the results of weighting the attributes of a therapy are also dependent on contextual parameters such as the disease, the patients and the perspective of the decision makers. For example, young patients with diabetes might feel most strongly about increasing or maintaining vitality whereas physicians with an eye on their patient’s long-term health might consider glucose control to be of primary importance.

In an alternative method for swing weighting the intrinsic values of a therapy’s attributes are plotted in a graph along with the values of these attributes in comparison to those of alternative
therapies (Figure 13).

If the attributes of the diabetes therapy that were scored in the previous example are colour coded and placed on the grid, it is then easy to discern those attributes with the highest intrinsic and contextual significance (Figure 14).

Dr Felli concluded by reiterating the importance of stakeholder verification of swing weighting results. In fact, a substantial portion of the value of evaluation methodologies such as swing weighting lies in the opportunity they afford to encourage critical conversations among stakeholders as to such issues as the differences in individual perspective, the choice of comparators and the appropriateness of scale in decision making.
Natalizumab benefit-risk assessment using the BRAT framework together with the application and visualisation of MCDA methodologies*

Dr Diana Hughes
Vice President, Worldwide Safety Strategy; Primary Care Business Unit Lead, Pfizer Inc, USA

The research presented by Dr Hughes was conducted as part of the Innovative Medicines Initiative Pharmacoeconomical Research on Outcomes of Therapeutics by a European ConsorTium (IMI PROTECT) project, which is a public-private multinational partnership of 32 partners including academics, regulators and pharmaceutical companies coordinated by the European Medicines Agency. The Consortium’s Work Program 5 (WP5) focuses on the integration and representation of benefit-risk. In wave 1 of WP5, four case studies were performed that evaluated various frameworks and quantitative methods for benefit-risk assessment of Raptiva (efalizumab), Ketek (telithromycin), Acomplia (rimonabant) and Tysabri (natalizumab). This presentation centres on the evaluation of natalizumab.

Dr Richard Nixon led the PROTECT team that evaluated the benefit-risk profile of natalizumab, using the six-step process developed by the Pharmaceutical Research and Manufacturers of America Benefit-Risk Assessment Team (PhRMA BRAT). This method defines the decision context and selects, organises, evaluates and displays relevant benefit-risk information, supported by an EXCEL-based tool.

Natalizumab was approved in 2004 by the FDA for the treatment of relapsing remitting multiple sclerosis. In 2005, the drug was removed from the market because of an associated incidence of progressive multifocal leucoencephalopathy (PML), a rare neurological disorder. Because of overwhelming patient demand, it was reintroduced in 2006 with strict risk minimisation requirements. In 2009, due to occurrence of further PML, the Committee for Medicinal Products for Human Use (CHMP) reassessed the PML risk of natalizumab and confirmed the current approval.

In Step 1 of the BRAT approach, the decision context must be determined. In this case the decision questions were should natalizumab have been given marketing approval at the time of first registration and should natalizumab be kept on the market given that increased episodes of PML were observed? Other elements of the decision context were the natalizumab indication, which is relapsing remitting multiple sclerosis; its comparators, which were Avonex (interferon beta1) and Copaxone (glatiramer acetate); its timeframe for treatment, which was two years; and the perspective of the decision maker, which was the European Medicines Agency (EMA).

In Step 2, the value tree for the benefits and risks associated with the treatment and its comparators and the outcomes by which they will be measured must be created. Target product profiles and risk management plans are appropriate sources for this step. The PROTECT evaluation team started with a comprehensive draft value tree for natalizumab, reducing it to contain the parameters that were most relevant to benefit-risk assessment though comprehensive discussion and expert knowledge of the comparators’ profiles. Value trees must be updated whenever new information on risks or benefits becomes available. The BRAT tool provides excellent support both for the creation of the value tree and for its updates.

In Step 3, source data must be identified, extracted and aggregated. For natalizumab, the publicly available data from three main clinical trials were used: natalizumab versus placebo; interferon beta1 versus placebo and glatiramer acetate versus placebo.
acetate versus placebo. Comparisons of active compounds were established via placebo calibration. For a more complex analysis, a full network meta-analysis could have been required.

In Step 4, a customisation framework must be created, updating the framework when new information becomes available and “tuning” the value tree by creating temporary displays of relevant data and concealing others according to the needs of the stakeholder audience.

In Step 5, the importance of the outcomes must be assessed. The use of a forest plot allows different orderings (ranking) of benefit and risk parameters using techniques such as value tree order or point estimation. Linear additive models, such as multi-criteria decision analysis (MCDA) allow the comparison of different outcomes measured on different scales. Although it would be inequitable to compare, for example, the rate of relapse of multiple sclerosis to the rate of flu-like reactions because of their different value to patients and clinicians, by considering how valuable a performance metric is in the range of what is possible, one can compare the values of the metrics instead of the metrics themselves. When using linear additive models with swing weighting, value functions measure within-outcome importance whilst swing weights measure between-outcome importance (Figure 15).

To assess the weights of the benefits and risks associated with natalizumab, the BRAT team ranked each outcome according to its importance and then asked how much more important it was to avoid the top-ranked event compared with the other outcomes. An example of a question that might be asked to assess the importance of an outcome:

Imagine a clinical trial of 1000 patients with 1 patient developing PML in the treatment arm.

How many patients would need to have the benefit of the prevention of Expanded Disability Status Scale progression for you to be indifferent about this harm caused by the treatment?

They then repeated this exercise for each parameter, moving bottom-up through the value tree and comparing the top-ranked outcomes from each category with the other items in that category. Finally, the top-ranked benefit was compared to the top-ranked risk and individual weights could be calculated for each outcome (Figure 16).

In addition to MCDA, other linear additive models include Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) and Analytic Hierarchy Process (AHP). MACBETH is similar to MCDA, except that it provides a different method to derive benefit and risk weights. In this approach, for each pair of parameters, evaluators first qualitatively assess how much more attractive it is to move from worst to best for outcome a versus moving from worst to best for outcome b and keeping everything else at the worst measure. After checking the consistency of their answers, users than compute initial estimations of weights with optimisation and finally refine the weights while maintaining consistency.

In AHP, weights are elicited by making pairwise comparisons of the importance between criteria, using 1 to 9 on a relative scale. Weight is then calculated by finding the dominant eigenvector of the corresponding matrix and values functions are computed in a similar manner. No consistency check is performed but a score is developed.

In Step Six key benefit-risk metrics must be displayed and interpreted. The BRAT tool provides two options for a quantitative overview on benefits and risks focussing on risk differences or ratios: a summary table or a forest plot.

However, because the PROTECT team elicited weights through MCDA swing weighting, it was possible to use graphic displays such as waterfall charts that illustrated the individual
**Benefit-risk evaluations afford a method to deconstruct and understand a problem, assess the main value drivers of a decision and communicate issues in a transparent, rational and consistent way. . .**

**Conclusions**

Although benefit-risk analysis is conceptually easy, it can be challenging to operationalise. In particular it can be difficult to define consistent criteria across decision options, find data matching these criteria and elicit value judgements. The BRAT framework is well suited to benefit-risk analysis, but a benefit-risk assessment does not necessarily provide an answer. Benefit-risk evaluations afford a method to deconstruct and understand a problem, assess the main value drivers of a decision and communicate issues in a transparent, rational and consistent way, while allowing sensitivity analyses around different stakeholder perspectives.

Editors Note: In January 2012, under its Unified Methods for Benefit-Risk Assessment (UMBRA) platform, CIRS assumed responsibility for the further advance of the BRAT initiative through integration with the BRASS approach.

* Based upon the presentation “A case study using the BRAT framework for BR Assessment” given on September 2012 by Richard Nixon, Modeling and Simulation, Novartis and Christoph Dierig, Global Integrated Analysis, Bayer Pharma

---

**Figure 17.** A waterfall diagram shows that even though PML was rated heavily for its importance, it made a relatively small contribution to the benefit-risk profile because of its rarity.

Two-way sensitivity analyses of the PML parameter and of the elicited weights in the natalizumab assessment were performed by the BRAT team and demonstrated that the incremental benefit-risk ratio that had been developed was robust in the face of these changes.

The length of each bar gives the contribution to the overall BR
- Denominator in the BR of one EDSS progression
- Green = positive BR
- Red = negative BR
- The contribution to the overall BR of PML is very small
Valuing options – A case study

Dr Marilyn Metcalf
Director, Benefit Risk Analysis USA GlaxoSmithKline, USA

The GSK Benefit-Risk Evaluation Group
The benefit-risk evaluation (BRE) group at GSK works to embed their approach for the measurement and articulation of the key benefits and risks of new medicines within product teams for their use at global safety board milestone reviews of new medicines and beyond. In this approach, teams are encouraged to use a value tree to facilitate framing the evaluation of the medicine in terms of the disease context, unmet need, the methods for measuring benefits and risks, the differences in sub-group effects and the reversibility and monitorability of adverse events as well as their long- and short-term occurrence.

Although evaluations are based on efficacy and safety, it should be understood that regulators, patients and other stakeholders will also need to know the medicine's likely effects on clinical outcomes and patient experience and its ultimate impact on disease. To enable the communication of those effects and to serve as a backdrop for informed discussion, a template was developed to create graphs which display the benefits and risks for the medicine side by side on a scale appropriate for the data.

A fictionalised case study
To illustrate the BRE process, Dr Metcalf presented a hypothetical case study that reflected multiple cases, fictionalised as an add-on therapy submitted for evaluation in the treatment of a chronic condition. The benefits associated with this medicine, which were improvements in a clinical endpoint, quality of life and function as well as symptom relief were measured with clinical test results, a disease-specific quality of life scale, observation and improved clinical measures, respectively. Potential adverse events included anaphylaxis, nausea, headache, anxiety, loss of appetite and drowsiness. In addition, a liver signal had been detected in animals that had not yet been observed in humans. Given the impact of Hy’s law, which according to the US Food and Drug Administration Guidance on Drug-Induced Liver Injury is "an observation that pure hepatocellular injury sufficient to cause hyperbilirubinaemia is an ominous indicator of the potential for a drug to cause serious liver injury," the signal detected in animals was regarded as a potential serious risk to be carefully evaluated.

Clinical trial endpoints were developed that were believed to be of importance to healthcare providers and patients based on key opinion leader feedback and the quality of life scale. As the evaluation discussion progressed, the benefits were listed in their relative order of importance as well as the level of evidence that had been accrued. Harms were included that were more worrisome to physicians and more disruptive to patients and that occurred in more than 2% of patients. Patient-reported outcomes were differentiated from those observed by healthcare professionals. (Figure 18).

For this fictitious product, benefit-risk evaluations were conducted at the time of proof of concept, submission and post-submission. Two types of graphs were created to display the different types of clinical trial measurements: treatment differences and odds ratios. Because the evaluated drug was an add-on-therapy to the existing standard of care, the standard of care was the comparator. Additional information about the graphs included:

- Benefits were measured on different scales and were therefore shown with different axes; "no difference" measures were aligned and there was some attempt to keep scales to comparable size among sections of the graph. Risk measures were reversed to move in the right direction versus the benefit measures.
- No summary measure of incremental net benefit or other overall "score" was used.
- Two populations with the same condition had been studied and rather than combine them, their measures were colour coded to show benefit in each group. This approach also works well when background rates are needed to provide context for benefit or risk.
- Differences among sources of data were noted. In this case, benefits were from 1-year and 2-year population phase 3 studies, while risks were from the primary safety population phase 3 studies and one phase 2 study.
Because the elements of a created value tree should be transferred to the graphs or their absence explained, it was necessary for product teams to report that there were no occurrences of liver signals applicable to Hy’s Law for this medicine (Figure 19).

Of note, although these methods were used to evaluate the benefit-risk profile of the drug relative to its comparator, it is possible to conduct the same evaluation of a medicine itself, as might be required for therapies for conditions with no available treatment. The same methods and discussions would apply regarding the nature and amount of benefit versus the nature and amount of risk associated with the medicine and how well that risk could be managed.

**Risk management**

In the risk management aspect of this evaluation, methods were sought to identify patients most at risk for anaphylaxis, which was judged to be the most worrisome of the potential adverse events associated with the drug. It was ensured that the team would include proper warnings to healthcare professionals and to patients around the signs and symptoms of anaphylaxis and the option of the delivery of the first dose in the presence of a healthcare provider was considered. The need for further understanding of the severity of associated headaches was noted and the team was advised to explore and address drowsiness issues, including the possible need for additional warnings around driving and the use of heavy machinery and whether bedtime dosing would be an alternative solution or if this might produce a “hangover” effect.

**Future plans**

Moving forward in the light of all of learnings, the BRE group plans to enhance elements of graphing benefit-risk by offering more choices and more opportunities for formal weighting. In addition, opportunities for quantification beyond graphing are being explored directly by the GSK Statistical Methods Benefit-Risk Working Group and through participation in external groups such as the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) and the Quantitative Sciences in the Pharmaceutical Industry (QSPI) Benefit-Risk Working Group. As part of the process of continuous improvement, the BRE group is also looking to incorporate patient perspectives into evaluations and to provide more step-by-step guidance and training in benefit-risk assessment and framing.

**Reference**

Structured assessment of benefit-risk: TYSABRI® (natalizumab) case study

Dr Carmen Bozic
Senior Vice President and Global Head, Safety and Benefit-Risk Management Biogen Idec Inc

There is high degree of organisational interest and support for structured benefit-risk assessment at Biogen Idec, where it is recognised that this process requires collaboration among many internal functions including safety, clinical development, biometrics and regulatory divisions. The basic benefit-risk framework used at Biogen Idec is consistent with the CIRS Unified Methodologies for Benefit-Risk Assessment (UMBRA) eight-step framework and employs a variety of methods including formal approaches to weighting, as well as clinical judgement.

Structured benefit-risk assessment has been used in regulatory filings for several Biogen Idec products and these assessments have entailed a variety of methodologies including those developed by the Pharmaceutical Research and Manufacturers Benefit-Risk Action Team (PhRMA BRAT; now administered through CIRS), conjoint analysis, Markov modelling, Number Needed to Treat/Number Needed to Harm measurement and the US FDA’s Benefit-risk table. Dr Bozic presented an example of the use of the PhRMA BRAT methodology in a regulatory appeal process for fampridine and of the use of Markov modelling and conjoint analysis in the analysis of the benefits and risks of natalizumab.

Fampridine

FAMPYRA (fampridine) is prolonged release 4-aminopyridine, which is a voltage-dependent potassium channel blocker approved in the EU, US and other countries for improvement in walking in patients with multiple sclerosis with walking disability. Although it was approved in the United States in 2010, fampridine initially received a negative opinion by the Committee for Medicinal products for Human Use (CHMP) because of questions surrounding the clinical meaningfulness of the primary outcome in clinical trials and because of concerns regarding the risk of seizure. Subsequently, this decision was appealed and the benefits and risks of fampridine were presented quantitatively using the PhRMA BRAT methodology, showing efficacy in both the primary and secondary endpoints and in the results of a post hoc analysis requested by the CHMP. The appeals presentation also included a demonstration of a slight risk of moderate-to-mild occurrence of non-serious adverse events. However, occurrence of seizure, the most serious potential adverse event, was shown to be higher in the placebo group than in those treated by fampridine. As a result of the appeals process, fampridine was approved by the CHMP in 2011.

Natalizumab

TYSABRI (natalizumab) is a monoclonal antibody against alpha-4 integrins approved in more than sixty countries for relapsing multiple sclerosis (MS). The drug has shown significant efficacy, demonstrating a 68% reduction in annualised relapse rate of MS and a 42% to 54% reduction in the progression of disability. However, natalizumab is also associated with risk of progressive multifocal leukoencephalopathy (PML) and a global risk management plan has been implemented since US and EU marketing approval in June 2006. In the US, all patients receiving natalizumab are enrolled in the Tysabri Outreach: Unified Commitment to Health (TOUCH) program, a risk management plan with elements to ensure safe use and in the EU and the rest of the world, the majority of patients receiving the drug are enrolled in registries or observational studies. At the time of natalizumab approval, the risk of PML was thought to be about 1 in 1,000 based on the clinical trial data.

Assessment of the benefits and risks of natalizumab has been ongoing since 2006, with efficacy data collected in addition to adverse event incidence. Information sources include new data from clinical trials and real-world effectiveness from registries and studies. Assessments of risk were made through healthcare outcomes such as PML survival, incidence and risk factors and PML risk stratification.

In a post hoc analysis of the AFFIRM (Natalizumab Safety and Efficacy in Relapsing Remitting Multiple Sclerosis) trial, it was demonstrated that 37% of patients who received natalizumab were free from disease activity over the course of the two-year study compared with 7% of patients receiving placebo. In addition, 69% more patients receiving natalizumab experienced a sustained improvement in physical disability compared with those receiving placebo and as measured by annualised relapse rates, a consistent effectiveness for the drug was observed across registries and observational...
From November 2004 to September 2012, 108,300 patients received natalizumab resulting in exposure of 224,718 patient years (Figure 21). Among those patients, 302 cases of PML were reported, 236 of whom were alive in September 2012, resulting in a 78% survival rate, compared with a 30% to 50% survival rate among patients with PML in other populations. This increased survival rate may be correlated with the education for and early diagnosis of PML that has resulted from natalizumab risk management programme. In addition, approximately two thirds of patients survived PML with mild-to-moderate disability as measured by Karnofsky performance status.

Analysis of data from global registries allowed the creation of a PML risk stratification algorithm using three risk factors for the increased occurrence of PML among patients receiving natalizumab: treatment duration, especially beyond 2 years; immunosuppressant use prior to receiving natalizumab and the presence of anti-John Cunningham virus (JCV) antibodies. Stratification of patients according to the algorithm reveals that patients who are anti-JCV antibody negative have a ≤ 0.09 in 1000 risk of contracting PML. Risk for patients who are anti-JCV antibody positive, can be further stratified according to prior antibody use and the duration of their exposure to natalizumab, resulting in a number of different risk levels, progressing to the highest risk of 11.1 in 1000 for patients who are anti-JCV antibody positive, have had prior antibody use and who have been exposed to natalizumab for between 25 and 48 months. Anecdotal evidence suggested that the acceptability of risk was high among patients receiving natalizumab, even among those in the highest risk category, who typically had few other treatment options. This evidence correlated with subsequent quantitative research in the form of Markov modelling and conjoint analysis.

Markov model
The objective of the Markov modelling for natalizumab was to compare long-term health changes, using quality of life years (QALY), for the natalizumab cohort with the health profiles modelled for a natural history cohort, a cohort treated with interferon β-1a and a cohort treated with a "perfect" MS treatment. In the analysis, net health gains and losses over 20 years associated with treatment over 20 years were calculated, with gains defined as delayed progression or reduced relapse rate and losses as non-PML side effects or PML.

An updated Markov model for natalizumab was also created using new data on increased incidence over time and improved survival outcomes for PML and the conclusions were the same. The Markov model results were submitted to EU regulators as part of a variation submission and was subsequently referenced in their assessment report.

Conjoint analysis
The objective of the benefit-risk conjoint analysis for natalizumab was to quantify the willingness of patients to accept treatment-related risks in exchange for improvements in outcomes. In this method, products described by various attribute levels are compared and from many comparisons it is possible to calculate overall preferences and preference functions and assuming a certain level of benefit and a maximum tolerated risk.

Data were collected via a trade-off questionnaire survey of stakeholders using a series of comparisons. The most typical stakeholders surveyed are patients, but the method can be used with other groups, such as prescribers. In this survey, 651 people with multiple sclerosis answered a series of choice format questions regarding hypothetical Treatment A and Treatment B, which were each associated with varying levels of efficacy and risks that included PML.

For example, given the following benefits and
potential harms, respondents were asked to select which treatment option they would prefer:

- With Treatment A you would experience 4 relapses during the next 5 years
- With Treatment B you would experience 0 relapses during the next 5 years
- With Treatment A your MS would get worse in 3 years
- With Treatment B your MS would get worse in 5 years
- With Treatment A 5 patients out of 1,000 may die or experience severe disability from PML within 10 years With Treatment B 0 patients out of 1,000 may die or experience severe disability from PML within 10 years

Results of the survey revealed that in return for a natalizumab-level efficacy, patients were willing to accept a 3.8 per 1,000 annual risk of death or severe disability due to PML or 7.6 per 1,000 risk over 2 years or 15.2 per 1,000 risk over 4 years.12 These data were also submitted to EU regulators as part of a variation submission.

Lessons learned and next steps

Although tools, models and processes assist in decision making, human judgement plays the most essential role.

Although tools, models and processes assist in decision making, human judgement plays the most essential role. In addition, although there are multiple available structured methods for benefit-risk assessment, no standard model has prevailed and regulators are not aligned on the best approach. Furthermore, choosing outcomes, that is, the benefits and risks to be evaluated, can be subjective. However, structured benefit-risk assessment may facilitate transparent discussions and decisions on benefit-risk both internally and with regulators and Biogen IDEC will continue to explore a variety of methods, including use of weighting, to facilitate internal decision making and to engage regulators in the review of their products.

References

**Weighting and valuing evidence in drug development and regulatory decision making: A situational analysis**

**Dr Becky Noel**  
*Senior Research Scientist, Eli Lilly, USA*

For a new drug or device to gain market access, it has to demonstrate compliance to manufacturing standards, an acceptable safety profile for its indication and a proven effect on specific efficacy parameters. Spanning these three pillars of quality, safety and efficacy, the therapy must also evince a positive benefit-risk balance to achieve marketing authorisation.

Multiple international initiatives over the past decade have helped to inform the current understanding of the topic including those of the EMA, the US FDA, the Pharmaceutical Research and Manufacturers of America (PhRMA), the Innovative Medicine Initiative (IMI), the International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) and CIRS Unified Methodologies for Benefit-Risk Assessment (UMBRA).

In just the past several years, a fairly significant body of work has been established through these initiatives:

- The last of five Work Projects of the EMA Benefit-Risk Methodology Project is currently in progress and the Pharmacovigilance Assessment Committee was established in 2012.
- The US FDA has initiated pilots of the use of their benefit-risk framework and the inaugural meeting of a series of 20 patient benefit-risk workshops took place in April 2013.
- The IMI Pharmacoepidemiological Research on Outcome of Therapeutics by a European Consortium (PROTECT) group is conducting a multi-year review and evaluation of methodologies used to model effects of medicines and the elucidation of patients’ preferences and integrating effects and preferences.
- As part of the E2C ICH guidance, an emphasis is being placed on integrated benefit-risk assessments and Periodic Benefit-Risk Evaluation Reports (PBRERs) have replaced periodic safety evaluation reports.
- The UMBRA initiative has advanced the pro forma developed by the Consortium for Benefit Risk Assessment (COBRA) and assumed responsibility for the further development of the work of the PhRMA Benefit-Risk Action Team (BRAT).

In addition, special interest groups and work streams dedicated to the topic of benefit-risk assessment have been formed in such organisations as the Drug Information Association (DIA), the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the European Federation of Technicians in the Pharmaceutical Industry (EFIT). All of these activities, especially newly legislated requirements for benefit-risk evaluation and the progress that has been made in the validation of the various methodologies, has established benefit-risk assessment as a prominent, global, multi-stakeholder topic.

Despite this progress, challenges remain, including those that surround the facilitation and implementation of the organisational changes that are necessary within both industry and regulatory agencies to promote the concept of benefit-risk assessment as well as to address the need to build capacity, experience and expertise in this area. Regulators want to ensure that progress is measured and thoughtful and that the position of clinical judgement is preserved. From the industry perspective, convergence, clarity and guidance around regulatory expectations are required. Both stakeholder groups are concerned that unstructured transparency in benefit-risk assessment may not provide the necessary context underpinning decision making (Figure 22).

**Weighing and valuing**

Because multiple initiatives have tested the operational and scientific validity of frameworks and methodologies and methodological approaches to weighting, such as those used in multi-criteria decision analysis are relatively well understood, it is reasonable to assume that methodology is not the only or even the primary impediment to weighting and valuing in decision making. Recognition is required that the contentiousness surrounding weighting cannot be addressed by focusing on the numbers alone. In addition to any methodological work that may be needed to advance the science of weighting, the policy and
operational requirements for weighting must also be developed.

Operational questions must be answered, such as whether to use weighting in all benefit-risk decisions or to reserve it for those evaluations with specific complexity or intricacy.

If weighting is to be performed, common criteria should be developed to inform the choice of weighting method used and the communication of best practices encouraged to build experience and expertise. Universal understanding of weighting among industry and regulatory agencies cannot be assumed and the value of continued intensive education using non-technical terminology to increase comprehension cannot be underestimated.

Data and societal norms are needed to inform weights to prevent them from being “guesstimates.” Decisions must be made as to whose value preferences take priority and to clearly define the methods for the elicitation of those preferences selected. Finally, it must be determined if stakeholders’ preferences are an a priori requirement for weighting or if tools should be provided to allow individual decision makers to explore their own preferences and their consequent decisions.

<table>
<thead>
<tr>
<th>Challenges and Concerns: Institutional B-RA</th>
<th>Industry</th>
<th>Regulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitating and promoting organizational change management</td>
<td>Facilitating and promoting organizational change management</td>
<td></td>
</tr>
<tr>
<td>B-R capacity and capacity building</td>
<td>B-R capacity and capacity building</td>
<td></td>
</tr>
<tr>
<td>Regulatory divergence, frequently expressed as “Which framework?”</td>
<td>Desire to preserve expert clinical judgment</td>
<td></td>
</tr>
<tr>
<td>Need for convergence and harmonization</td>
<td>Need to bring along stakeholders, have structured public engagement, at an appropriate speed (e.g., risk of moving too quickly)</td>
<td></td>
</tr>
<tr>
<td>Unstructured transparency: context is critical</td>
<td>Transparency associated with characterizing uncertainties, value judgments and trade-offs</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix: Workshop Attendees

<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Responsibilities</th>
<th>Organization/Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Cristina Alonso</td>
<td>Head, Global Regulatory Affairs, Latin America</td>
<td>Bayer Healthcare Pharmaceuticals Inc, USA</td>
</tr>
<tr>
<td>Dr Gary Bloomgren</td>
<td>Vice President, Safety and Benefit-Risk Management</td>
<td>Biogen Idec, USA</td>
</tr>
<tr>
<td>Dr Carmen Bozic</td>
<td>Senior Vice President and Global Head, Drug Safety and Benefit-Risk Management</td>
<td>Biogen Idec, USA</td>
</tr>
<tr>
<td>Dr John Bridges</td>
<td>Associate Professor, Director of the MHS in Health Economics, Department of Health Policy &amp; Management</td>
<td>Johns Hopkins Bloomberg School of Public Health, USA</td>
</tr>
<tr>
<td>Dr Nadine Cohen</td>
<td>Senior Vice President, Regulatory Affairs</td>
<td>Biogen Idec, USA</td>
</tr>
<tr>
<td>Dr Bryan Dirks</td>
<td>Clinical Medicine Director</td>
<td>Shire, USA</td>
</tr>
<tr>
<td>Dr James Felli</td>
<td>Research Fellow</td>
<td>Eli Lilly and Company, USA</td>
</tr>
<tr>
<td>Dr Christine Hallgreen</td>
<td>PostDoc</td>
<td>Novo Nordisk A/S, Denmark</td>
</tr>
<tr>
<td>Dr Richard Hermann</td>
<td>Safety Science Physician</td>
<td>AstraZeneca, USA</td>
</tr>
<tr>
<td>Dr Diana Hughes</td>
<td>Vice President, Worldwide Safety Strategy, Primary Care Business Unit Lead</td>
<td>Pfizer Inc, USA</td>
</tr>
<tr>
<td>Dr Qi Jiang</td>
<td>Executive Director, Global Biostatistical Science</td>
<td>Amgen Inc, USA</td>
</tr>
<tr>
<td>Judy Kannenberg</td>
<td>Director, Regulatory Affairs</td>
<td>Astellas Pharma Global Development Inc, USA</td>
</tr>
<tr>
<td>Dr Haley Kaplowitz</td>
<td>Senior Director, Epidemiology</td>
<td>Allergan, USA</td>
</tr>
<tr>
<td>Beth Ann Knapp</td>
<td>Director, Regulatory Affairs Strategy, Cardiovascular</td>
<td>Takeda Global Research and Development Center, USA</td>
</tr>
<tr>
<td>Dr Elias Kouchakji</td>
<td>Executive Medical Director</td>
<td>Amgen, USA</td>
</tr>
<tr>
<td>Dr Bennett Levitan</td>
<td>Director, Epidemiology</td>
<td>Janssen Research &amp; Development, USA</td>
</tr>
<tr>
<td>Dr Meni Melek</td>
<td>Head, Hematology, Global Regulatory Affairs</td>
<td>Bayer Healthcare, USA</td>
</tr>
<tr>
<td>Dr Marilyn Metcalf</td>
<td>Senior Director, Benefit Risk Evaluation</td>
<td>GlaxoSmithKline, USA</td>
</tr>
<tr>
<td>Dr Steven Miller</td>
<td>Vice President, Regulatory Affairs</td>
<td>Janssen Research and Development, USA</td>
</tr>
<tr>
<td>Dr Filip Mussen</td>
<td>Head, Global Labeling Center of Excellence</td>
<td>Janssen Research and Development, Belgium</td>
</tr>
<tr>
<td>Dr Becky Noel</td>
<td>Senior Research Scientist</td>
<td>Eli Lilly and Company, USA</td>
</tr>
<tr>
<td>Kinnari Patel</td>
<td>Associate Director, Global Regulatory Sciences, US</td>
<td>Bristol-Myers Squibb Company, USA</td>
</tr>
<tr>
<td>Michelle Peralta</td>
<td>Leader of PV Risk Management, Americas</td>
<td>Takeda Development Center, USA</td>
</tr>
<tr>
<td>Dr Cynthia Richards</td>
<td>Senior Clinical Medical Director</td>
<td>Shire, USA</td>
</tr>
<tr>
<td>Dr Wendy Sanhai</td>
<td>Senior Director, Global Regulatory Policy and Advocacy</td>
<td>GlaxoSmithKline, USA</td>
</tr>
<tr>
<td>Dr Paul Seligman</td>
<td>Executive Director – US Regulatory Policy</td>
<td>Amgen Inc, USA</td>
</tr>
<tr>
<td>Dr Meredith Smith</td>
<td>Senior Scientific Director, Risk Management, Global Research and Development</td>
<td>Abbott Laboratories, USA</td>
</tr>
<tr>
<td>Dr Mark Taisey</td>
<td>President, Global Regulatory Affairs</td>
<td>Eisai Product Creation Systems, USA</td>
</tr>
<tr>
<td>Centre for Innovation in Regulatory Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patricia Connelly</td>
<td>Manager, Communications</td>
<td></td>
</tr>
<tr>
<td>Art Gertel</td>
<td>Senior Research Fellow</td>
<td></td>
</tr>
<tr>
<td>Lawrence Liberti</td>
<td>Executive Director</td>
<td></td>
</tr>
<tr>
<td>Dr Neil McAulane</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Professor Stuart Walker</td>
<td>Founder</td>
<td></td>
</tr>
</tbody>
</table>

BENEFIT-RISK FRAMEWORK: VALUING THE OPTIONS; 13 December 2012; Philadelphia USA