Focus – Benefit–risk assessment and frameworks

Evaluating benefit–risk during and beyond drug development: An industry view

Introduction
Benefit–risk evaluations of drugs have been conducted since the introduction of modern regulatory systems in the 1960s, following the thalidomide disaster. However, it has only been in the past decade that both industry and regulators have started to focus on the actual methodology for conducting such benefit–risk evaluations. Since then, many initiatives by regulatory and industry groups have emerged on this topic, as well as several methodologies to conduct such assessments.

Overview of regulatory and industry benefit–risk initiatives
On the regulatory side, the European Medicines Agency (EMA) was one of the first agencies to publicly discuss benefit–risk methodology. This was followed by enhanced benefit–risk guidance for clinical assessors and by the EMA’s Benefit–Risk Methodology Project, which had as its main objective to develop and test tools and processes for balancing multiple benefits and risks.1 Currently, four tools have been proposed within the scope of this project: a generic decision-making approach entitled the PrOACT-URL Framework, an Effects Table summary of benefit and risk outcomes, Multi-Criteria Decision Analysis (MCDA) modelling,2,3 and graphical displays. In the US, the FDA has started to pilot a “grid”-based framework (see Table 1) which includes key benefit–risk attributes. The FDA has also agreed with industry to include within the Prescription Drug User Fee Act (PDUFA V) reauthorisation proposal a five-year project to further develop and implement structured benefit–risk assessment in the drug approval process.4 In addition, the FDA has recently published a benefit–risk guidance for medical device premarket.5

On the industry side, a project that has received some attention is the Benefit–Risk Action Team (BRAT) Framework developed by the US industry association, Pharmaceutical Research and Manufacturers of America (PhRMA).6,7 It is a set of processes and tools for selecting, organising, summarising and interpreting evidence for benefit–risk decisions. It consists of six steps and includes a number of visual displays and recommended reports for transparent decision-making (see Figure 1). Several companies are currently piloting these BRAT tools for selected development compounds and marketed products.

The BRAT Framework has recently been moved to the Centre for Innovation in Regulatory Science for continued development. The Centre has also been sponsoring a benefit–risk methodology project entitled the CASS project, named after the four regulatory authorities included: Canada, Australia, Singapore and Switzerland.8 Finally, another project which involves not only industry but also European regulators and academics is IMI PROTECT, which focuses on benefit–risk methodology in the post-approval space.9

Today, there are no pharmaceutical regulatory or industry standards for the use of benefit–risk methodologies, nor a set of preferred tools, though recent draft EMA guidances on periodic safety update reports (PSURs)10 and the recently expanded template for rapporteur marketing authorisation assessment11 are excellent steps in that direction. There are, however, a number of common elements in most of the above methods, which are reviewed further below. It is also important to note that different methods may suit different purposes, ranging from informing late stage drug development to benefit–risk re-evaluation in the post-marketing space, and from a discussion and decisions aid to a vehicle to communicate such decisions.

Overview of benefit–risk tools
Although there is no agreed taxonomy for benefit–risk tools, one useful classification is qualitative frameworks, semi-quantitative frameworks, and quantitative models (see Figure 2). Qualitative frameworks are generally templates, grids or more visual displays that list key benefit and risk attributes and may summarise the key aspects of these attributes. An example component of such a framework is the value tree, a hierarchic graphic representation of a set of attributes, originally developed within MCDA and used in BRAT and CASS (see Figure 3). Semi-quantitative models are usually built on a qualitative framework but include tabular or graphical tools to display and summarise the metrics associated with the key benefit and risk attributes. The BRAT framework uses these kinds of visual tools in the format of a key benefit–risk summary table (see Table 2) and with risk or rate difference forest plots (see Figure 4). Quantitative models typically calculate a benefit–risk score with associated uncertainty, and thus require a method to put benefits and risks on the same scale. They usually include an algorithm to

Authors

Keywords
Benefit–risk; Evaluation; Methodology; Benefit–Risk Action Team (BRAT) framework; Qualitative/semi-quantitative/quantitative frameworks; Forest plot; Measures; Outcomes.

Abstract
In the past decade, both industry and regulators have begun a significant number of initiatives to measure benefit–risk in a systematic, methodical manner. This article reviews these initiatives and summarises some of the more common approaches used in benefit–risk assessment.
combine metrics for specific benefit and risk outcomes with measures that quantify the clinical impact, or weight, of an attribute. An MCDA model and stated choice preference studies are typical methods used in such a quantitative approach.

There is today a growing belief that a single approach for benefit–risk evaluations is not ideal, but that instead a toolbox should be made available that can be applied depending on the complexity of the data, the complexity of the benefit–risk profile, and other factors. Such a toolbox would consist of a qualitative framework supplemented with visual and quantitative methods, with the understanding that the more complex (typically weighted) approaches have semi-quantitative approaches as a necessary precursor.

What structured processes do companies use, and for what purpose? During drug development: While all steps in structured approaches to benefit–risk assessment can be important, during drug development the more qualitative steps provide considerable value and are the foundation for the use of semi-quantitative or quantitative methods during and beyond the regulatory approval phase. For example, referring to Figure 1, the decision context (e.g., medical need, treatments being compared, indication, population, audience for analysis) needs to be carefully defined and is typically based on application of the clinical studies for health authority approval. Moreover, identifying the outcomes (step 2) can be of paramount importance. This activity generally has two components: (i) identifying outcomes for benefit–risk assessment using a value tree or other exercise and (ii) developing precise measures for these outcomes that work in a benefit–risk context.

(i) Identifying outcomes. Clinical trials include a statistical analysis plan (SAP) that rigorously defines all efficacy and safety endpoints to be analysed. These endpoints generally work very well for understanding efficacy and safety separately; however, they are not always sufficient for benefit–risk assessment. For example, it is quite common to encounter endpoints that are causally dependent or even double-count the same event. In a typical cardiovascular trial for an anticoagulant, the primary efficacy outcome includes cardiovascular death and stroke, and the primary safety outcome includes some definition of moderate and severe bleeding. Hemorrhagic strokes and fatal bleeds will be counted twice – once in the primary efficacy outcome and second in the primary safety outcome. Any benefit–risk assessment based on comparing the primary endpoints is potentially misleading with this double-counting. A popular approach to ensuring that all relevant endpoints are accounted for in a benefit–risk assessment is to develop the value tree (see Figure 3). The value tree is meant to include only those outcomes that the sponsor/decision makers regard as relevant to a benefit–risk assessment.

### Table 1: FDA benefit–risk framework

<table>
<thead>
<tr>
<th>Decision factor</th>
<th>Evidence and uncertainties</th>
<th>Conclusions and reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of condition</td>
<td>Summary of evidence:</td>
<td>Conclusions (implications for decision):</td>
</tr>
<tr>
<td>Unmet medical need</td>
<td>Summary of evidence:</td>
<td>Conclusions (implications for decision):</td>
</tr>
<tr>
<td>Clinical benefit</td>
<td>Summary of evidence:</td>
<td>Conclusions (implications for decision):</td>
</tr>
<tr>
<td>Risk</td>
<td>Summary of evidence:</td>
<td>Conclusions (implications for decision):</td>
</tr>
<tr>
<td>Risk management</td>
<td>Summary of evidence:</td>
<td>Conclusions (implications for decision):</td>
</tr>
</tbody>
</table>

### Figure 1: Steps in the BRAT Framework

- Define decision context
- Identify outcomes
- Identify and extract source data
- Customise framework
- Assess outcome importance
- Display and interpret key benefit–risk metrics
- Decision and communication of benefit–risk assessment
excessively numerous, do not double-count events, and separate outcomes with different patient experiences. This exercise also provides the opportunity to consider whether it is possible to simplify the assessment by defining one composite efficacy endpoint and one composite safety endpoint that can be compared, a popular simplification in the benefit–risk literature, or a single composite endpoint that includes both benefit and risk events of comparable clinical significance.

(ii) Defining measures. Measures for many of the outcomes in the value tree can be taken from an SAP. However, there are numerous issues unique for benefit–risk assessment, for example:

- One common issue results from the convention that, in many clinical trials, efficacy outcomes are defined over an intent-to-treat (ITT) population and time period, while safety outcomes are defined over a safety population and time period. Typically, ITT includes all patients and events from randomisation through to follow-up visits, while the safety population includes all patients and events from first dose to a few days or a follow-up period after last dose. Having efficacy and safety measures defined over different populations and over different time periods can be very problematic when comparing benefits and risks, particularly in the important cases when there is a large lag between randomisation and first dose or when important events occur in the post-treatment period (eg, rebound phenomena).

- Another critical concern for measures in benefit–risk is whether they count events or patients. Outcomes such as death and disabling stroke inherently count only one event per patient, but events such as bleeding, migraine headaches and joint pain can occur repeatedly to a patient during a study. When considering the clinical impact of events in a benefit–risk balance, this distinction is critical, as a single bleeding event or single migraine can occur repeatedly to a patient during a study. When considering the clinical impact of events in a benefit–risk balance, this distinction is critical, as a single bleeding event or single migraine are completely different experiences than having many during a year.

- A third concern arises with composite endpoints. Composite endpoints combine several events under one heading, typically done to provide greater statistical power for a set of related endpoints or to simplify interpretation of a complex set of events. There are several potential complications when composite endpoints are used in benefit–risk: (i) the events combined together may span a wide range of clinical impact, (ii) a composite can obscure cases where some of the component events favour one treatment and other components events favour the other treatment, (iii) composite endpoints are often measured with a time-to-first-event analysis. With this approach, patients who experience multiple events only contribute one event to the analysis. For example, for a composite of “death, myocardial infarction (MI) or stroke,” a patient who first has an MI, then a stroke,
From an industry perspective, it is believed that in clinical overviews and in CHMP hearings there would be great value in better articulating the benefit–risk profile and in having a more systematic discussion of relevant benefit and risk aspects and their associated uncertainties. This would enhance a rational decision-making process and improve transparency of the decision for all parties. At present, however, there are few regulatory and industry guidelines for benefit–risk assessment, though the EMA guidance document on the content of the day 80 critical assessment report and recent guidelines for PSURs provide very valuable starting points. The critical assessment report guidance stresses the importance of listing the benefits and risks together, how they are measured, point estimates and confidence intervals for the effect size vs the comparator, sources of uncertainty on these values and related information. This section is then followed by a qualitative description of the clinical relevance or weight of each benefit and risk, as assessed by a rapporteur or co-rapporteur, followed by an explanation of how the combined favourable effects are judged to exceed (or fail to exceed) the combined unfavourable effects. In particular, these guidelines stress the importance of distinguishing between data and regulator value judgments.

More pragmatically, a benefit–risk assessment typically starts with a reiteration of medical need and a summary of relevant benefit and risk endpoints results. In general, a table or figure (eg, effects table or key benefit–risk summary table) showing the data for both benefit and risk endpoints simultaneously is very valuable (see Figure 3). If this table includes point estimates for each treatment studied and some measure of difference between the treatments, such as risk or rate differences, clinical judgment is often sufficient to enable decision-makers to balance the benefits and risks and make a decision from this table alone. For dichotomous endpoints, forest plots can also be very helpful in communicating data on multiple benefits and risks simultaneously (see Figure 4). These displays have the advantage of enabling reviewers to use their own judgment to decide which endpoints to compare against one another for benefit–risk balance. Number-needed-to-treat (NNT) and number-needed-to-harm (NNH) are also sometimes used to summarise differences between two treatments. For a small number of endpoints, NNT and NNH can be a useful adjunct to these tables, although there may be mathematical challenges in estimating and interpreting confidence intervals around these quantities. Finally, to assess benefit–risk over the course of treatment, it can be useful to display Kaplan-Meier curves of a key benefit and key risk in the same plot. Overall, it is difficult to overstate the value of clear, concise visualisations for communicating benefit–risk assessments in a regulatory context.

### Table 2: Example of key benefit–risk summary table

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study drug risk (/1,000 pts)</th>
<th>Comparator risk (/1,000 pts)</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>Rapid onset 271</td>
<td>248</td>
<td>1.13 (1.00, 1.27)</td>
</tr>
<tr>
<td></td>
<td>Headache relief 643</td>
<td>633</td>
<td>1.04 (0.94, 1.15)</td>
</tr>
<tr>
<td></td>
<td>Pain-free response 383</td>
<td>349</td>
<td>1.16 (1.03, 1.30)</td>
</tr>
<tr>
<td></td>
<td>Sustained response 285</td>
<td>295</td>
<td>0.95 (0.80, 1.14)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Reduced sensitivity to sound and light 530</td>
<td>505</td>
<td>1.10 (0.94, 1.30)</td>
</tr>
<tr>
<td>Other</td>
<td>Reduction in functional disability 540</td>
<td>480</td>
<td>1.28 (1.09, 1.49)</td>
</tr>
<tr>
<td></td>
<td>Reduction in nausea or vomiting 604</td>
<td>517</td>
<td>1.43 (1.22, 1.67)</td>
</tr>
<tr>
<td><strong>Risks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual risks</td>
<td>Transient triptans sensations 43</td>
<td>52</td>
<td>0.83 (0.61, 1.14)</td>
</tr>
<tr>
<td></td>
<td>Central nervous system adverse events 53</td>
<td>45</td>
<td>1.18 (0.92, 1.51)</td>
</tr>
<tr>
<td></td>
<td>“Chest-related” adverse events 58</td>
<td>21</td>
<td>2.93 (2.04, 4.20)</td>
</tr>
</tbody>
</table>
Focus – Benefit–risk assessment and frameworks

Post-approval: While most of the benefit–risk approaches described for development and health authority meetings apply in the post-approval setting, there are several critical differences, including:

- Referring to Figure 1, it can be much more difficult to define the appropriate decision context. What is the appropriate comparator – the comparators used in the clinical studies used to approve the drug in question, the current standard of care, the best in class? Should the population be that used to approve the drug or the more heterogeneous population for which the treatment is actually used? Should the usage be as in the drug label or as in practice, when those are not aligned?

- Post-approval benefit–risk assessment generally requires the use of observational data. There are often several observational databases to consider, and generally no single database will meet all the desired requirements. For example, endpoints from clinical trials are not always captured in observational databases, or are defined differently, and benefit data (especially patient-reported outcomes) are typically difficult to obtain in observational data. Identifying and defining the relevant and usually very comprehensive efficacy and safety data set may be a major challenge. One solution may be through the use of registries or large simple trials.

- The framework or quantitative method used for benefit–risk assessment during approval can often be applied with updated post-marketing data. However, if a new type of risk is identified from post-approval use, or if there is a substantial change in the clinical impact of some of the benefits or risks, such as may result from changes in treatment paradigms, then the method may need to be revised.

Implications and conclusions

Each kind of decision approach has advantages and disadvantages and may therefore not be easily applicable in every decision setting. Quantitative methods that combine metrics with weights may, if well developed, lead to a very high quality benefit–risk decision, but require a major effort and resources, and have less transparency. They use for regulatory review is therefore less appropriate and should be reserved for exceptional situations for drugs with a complex benefit–risk profile. In addition, there remain a number of challenges with quantifying clinical expert judgments in terms of weights, and there is understandable scepticism about using very refined weighting scales as they may provide a false sense of accuracy. Qualitative or semi-quantitative methods seem appropriate for most benefit–risk decisions, and can be more easily used by industry, especially in regulatory settings such as CHMP oral hearings or FDA advisory committee meetings.

One major area that requires further development and discussion is uncertainty. While statistical uncertainty is captured well in most decision approaches, work remains to be done with regard to better articulating the consequences of any gaps in the efficacy and safety data (e.g., dropouts) and the level of evidence available on the benefit–risk profile. During the past decade, very valuable new concepts and approaches for benefit–risk evaluations have emerged, and it is believed these kinds of approaches will ultimately lead to more systematic, predictable and transparent benefit–risk decisions. A collaborative effort between industry and regulators will be required to continue to advance the science of benefit–risk methodology, since, as we have argued above, there is no single or simple approach that would address all benefit–risk assessments. Eventually, we expect a set of common principles, standards and a toolbox of methods will emerge. Nevertheless, our experience and a considerable body of literature suggests that
the most useful aspect of a systematic benefit-risk approach will continue to be a qualitative structured framework, because of the transparency such an approach brings to the assessment. We encourage both industry and regulators to continue to discuss, apply and experiment with benefit-risk methodology in an open and collaborative manner, as we believe this will be key to the ultimate goal of a common and global benefit-risk platform, possibly within the scope of the International Conference on Harmonisation or similar global organisation.

Disclaimer
The opinions expressed in this paper are those of the authors and not necessarily the views of Janssen Pharmaceutical Companies of Johnson & Johnson.

References


11 Guidance document on the content of the rapporteur day 80 critical assessment report, 2011.


