REVIEW

BRITISH PHARMACOLOGICAL SOCIETY

Advancing structured decision-making in drug regulation at the FDA and EMA

Aris Angelis [ORCID] | Lawrence D. Phillips

London School of Economics and Political Science, London, UK

Correspondence
Aris Angelis, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.
Email: a.n.angelis@lse.ac.uk

The recent benefit–risk framework (BRF) developed by the Food and Drug Administration (FDA) is intended to improve the clarity and consistency in communicating the reasoning behind the FDA's decisions, acting as an important advancement in US drug regulation. In the PDUFA VI implementation plan, the FDA states that it will continue to explore more structured or quantitative decision analysis approaches; however, it restricts their use within the current BRF that is purely qualitative. By contrast, European regulators and researchers have been long exploring the use of quantitative decision analysis approaches for evaluating drug benefit–risk balance. In this paper, we show how quantitative modelling, backed by decision theory, could complement and extend the FDA's BRF to better support the appraisal of evidence and improve decision outcomes. After providing relevant scientific definitions for benefit–risk assessment and describing the FDA and European Medicines Agency (EMA) frameworks, we explain the components of and differences between qualitative and quantitative approaches. We present lessons learned from the EMA experience with the use of quantitative modelling and we provide evidence of its benefits, illustrated by a real case study that helped to resolve differences of judgements among EMA regulators.

KEYWORDS
benefit–risk assessment, decision analysis, drug regulation, EMA, FDA, MCDA

1 | BACKGROUND

Drug regulators decide whether a new medical product can be granted marketing authorization by assessing the drug's benefit–risk balance, that is, "balancing the desired effects or 'benefits' of a medicine against its undesired effects or 'risks'".[1] This is often a challenging and complex task given the extensive body of evidence submitted by the applicant of a New Drug Application (NDA) or Biologics Licensing Application (BLA). A number of additional considerations going beyond the typical measures of a drug's benefits and risks can further complicate the evaluation procedure, including the severity of the condition, the unmet clinical need based on the availability of current therapies, the uncertainty about how available clinical trial data might translate to broader use after approval, and the potential need for risk management tools.[2,3]

Explicit value judgements on prescription drugs' effects and their trade-offs are necessary to understand their benefit–risk balance. Although clinical trials provide scientifically *objective* data about safety, efficacy and quality, regulators make *subjective* judgements about what evidence concerns the intended human use of the drug, how to evaluate the safety and efficacy data, how clinically relevant the evidence is, how relatively important they are, and other

considerations. In short, the benefit–risk balance of a drug depends on both *objective* clinical data and its *subjective* interpretation.

The FDA's benefit–risk framework (BRF) is an important advancement in the drug regulatory landscape.[3] Following its earliest conception,[4] the FDA has been legally committed to its phased implementation for use within the regulatory review process and documentation of NDAs and BLAs in accordance with the Prescription Drug User Fee Act (PDUFA) V in 2012 and PDUFA VI in 2017.[3] The goal of the FDA's BRF is to "improve the clarity and consistency in communicating the reasoning behind drug regulatory decisions", helping drug sponsors to better understand the factors that contribute to decisions but also ensuring that FDA assessments are readily understood.[5]

Although in the PDUFA VI implementation plan the FDA states that it will continue to explore "more structured or quantitative decision analysis approaches, methods, and tools", it also restricts their use within the current, purely qualitative framework, not leaving room for a possible extension to a fully quantitative method.[3] By contrast, European regulators and researchers have been exploring a role for quantifying the benefit–risk balance as part of the approval processes, through "explicit" decision-making approaches involving well-defined evaluation criteria, their relative importance, and data relating to product performance.[6,7] Overall, in contrast to quantitative decision analysis, qualitative approaches do not allow for the quantification of values, uncertainties and trade-offs, nor their aggregation. Based on cognitive psychology, behavioural economics and decision science literature, the limited capacity of the human brain to aggregate multiple pieces of information can be obviated by assigning that task to quantitative models so computers can combine the information.[8] Numbers can be combined in ways that are impossible with words.

Such a quantitative approach could allow the FDA's BRF to support the appraisal of evidence, decision outcomes, and value communication that are vital to physician and patient decision-making. It could also facilitate regulatory decision-making and communication in complex decision contexts where many favourable and unfavourable effects of unequal relative clinical importance make it difficult to agree about the overall benefit–safety balance, or in cases surrounded by disagreement over evidence interpretation; for example, when surrogate measures are used as the basis of clinical benefit in place of actual clinical outcomes,[9] including for the case of advanced therapy medicinal products (ATMPs).[10] A real example of how a quantitative decision analysis approach facilitated a regulatory decision context characterized by multiple favourable and unfavourable effects under consideration is described in the relevant case study section below, for belimumab's (Benlysta) EMA review of systemic lupus erythematosus (SLE). An example of a problematic decision context that could be facilitated by quantitative modelling would be eteplirsen's (Exondys 51) FDA review, characterized by strong disagreement on evidence interpretation between FDA experts and staff.[11] Finally, a recent ATMP case associated with uncertainty in clinical benefit due to the use of surrogate endpoint would be voretigene neparvovec (Luxturna) for which current evidence failed to support a curative benefit for most patients, with large heterogeneity in response rate and possible shorter duration of benefits than expected.[12]

Furthermore, a number of ongoing FDA activities represent additional opportunities for quantitative decision analysis methods to be used for evidence interpretation and aggregation, further strengthening the basis for their use. Such an approach could allow the exploration of patients' priorities and preferences on decisions, including helping to address controversies about the integration of subjective patient experience in the review process,[13] as envisioned by US legislators in the 21st Century Cures Act.[5] Another initiative as part of PDUFA VI would be the Model-Informed Drug Development (MIDD) Pilot Program, involving a variety of quantitative methods for balancing the risks and benefits of drugs in development to improve clinical studies' efficiency and regulatory success, the output of which could be leveraged.[14] Finally, following the Cures Act, real-world evidence (RWE) plays an increasing role in health care decisions, including for monitoring of post market safety for regulatory decisions and supporting innovative clinical trials designs, which could also be used to inform gaps in data.[15]

In this paper, we show how quantitative modelling, together with decision science, could complement and extend the FDA's qualitative BRF and improve decision-making by the EMA's Committee for Medicinal Products for Human Use (CHMP). In doing so, we are using evidence from the literature to highlight the limitations of qualitative approaches while referring to the EMA experience on quantitative decision analysis to showcase its advantages.

2 | SCIENTIFIC DEFINITIONS AND DRUG REGULATORY FRAMEWORKS FOR BENEFIT–RISK ASSESSMENT

Past interviews with over 50 drug experts in six European regulatory agencies revealed substantial differences of opinion about the meaning of benefits and risks.[16] Apart from incompatible meanings, many respondents failed to distinguish between the magnitude of an effect and the uncertainty of experiencing the effect. That confusion led the European Medicines Agency (EMA) to adopt in 2009 the 2×2 matrix shown in Figure 1.

These definitions are elaborated by EMA regulators, who use them in describing the benefit–risk balance of a new prescription drug.[17] Briefly, any patient benefit from taking a drug is a favourable effect, which includes the reduction or elimination of an undesirable symptom. Any side effects attributable to the drug taken are unfavourable effects. Uncertainty attends both kinds of effects. All four cells contribute to the benefit–risk balance. No matter how a drug's effects and uncertainty are measured, as they must be in making regulatory decisions to approve or clinical decisions to prescribe, judgements are required to interpret the evidence for its clinical relevance. Performance measures of a drug and its comparator might show a statistically significant difference, but that difference could be clinically weak or unimportant. Thus, both objective measures and the

Favourable Effects	Uncertainty of favourable effects
Unfavourable Effects	Uncertainty of unfavourable effects

FIGURE 1 The EMA matrix

subjective judgements of clinical value associated with the evidence, when considered alongside uncertainty about the effects, define the benefit–risk balance. Current practice is embodied in the FDA's BRF and the EMA's 80-Day Guidance documents.[18] The FDA's BRF is "a structured, qualitative approach focused on identifying and clearly communicating key issues, evidence and uncertainties in FDA's benefit–risk assessment and how those considerations inform regulatory decisions".[3] It is composed of two key elements: first, the Benefit–Risk Dimensions portion, outlining the critical elements that are considered in the BRF (analysis of condition, current treatment options, benefit, risk, and risk management), along with statements of "evidence and uncertainties" and "conclusion and reasons", and second, the Benefit–Risk Integrated Assessment, combining all dimensions in an overall analysis and providing an explanation or rationale of the regulatory recommendation/decision.

By contrast, since 2015 the EMA requests the assembly of an extended Effects Table (ET) for initial applications of new active substances and important extensions of indication applications, essentially a matrix of the drug's performance across those effects that are relevant for the purpose of licensing.[17] The purpose of the ET is to "improve consistency, transparency, and communication of the benefit–risk assessment"; it does so by summarizing only those favourable and unfavourable effects measured in the clinical trials for the drug alternative(s) and comparator(s) that were taken into account by the regulator, along with descriptions of their uncertainties. This provides the factual basis for discussions and value judgements by the regulators of the clinical relevance of the data and their interpretation, leading to the judgement of the overall benefit–risk balance.[18] More precisely, the ET contains the following information: names and definitions of each effect that is relevant to the overall assessment, the unit of measurement for each effect, data summarizing for all effects the key outcomes of the testing intervention and comparators, a statement of the strength of evidence and any major uncertainties or limitation for each effect, with references identifying the relevant part of the text or specific sources of data.

Using a common glossary for these terms along with the ET improves decision-making transparency and value communication.

3 | QUALITATIVE, SEMI-QUANTITATIVE AND FULLY QUANTITATIVE BENEFIT–RISK ASSESSMENT

For decision contexts in which a drug's benefit–risk balance is straightforward, the FDA's current qualitative BRF likely serves the needs of assessors. However, for more complex decisions that involve greater uncertainty about the benefit–risk balance, quantitative approaches using decision analysis can provide useful complementary information. A relevant past example of a quantitative benefit–risk approach is the risk analysis conducted by the FDA's Center for Biologics Evaluation and Research (CBER) for the review of a home-use HIV test kit, which applied probability distributions and sensitivity analyses to model uncertainty of outcomes.[19] Another example is the decision analysis conducted by the FDA's Center for Devices and Radiological Health (CDRH) for weighing the benefits and risks of retrievable filters in patients without pulmonary embolism.[20]

Four fundamental components are shared for any benefit–risk assessment: the definition of Objectives and Criteria, the identification of Alternatives and Options, the collection of Data and Evidence, and the elicitation of Consequences and Preferences. Taking into account only these components reflects a purely *qualitative* approach. The approach becomes *semi-quantitative* if utilities or values, uncertainty, and trade-offs are quantified. Combining all these components using an aggregation model as an algorithm defines a *fully quantitative* approach.[7] These elements are illustrated in Figure 2.

For simpler cases, presenting only one or two favourable effects and whose adverse effects are minor, qualitative approaches may be sufficient. With more effects, but less than 6–10, a semi-quantitative approach, accompanied by an ET will ensure that all effects are considered. For more effects, and when there is disagreement among the assessors, particularly where trade-offs are involved, sensitivity
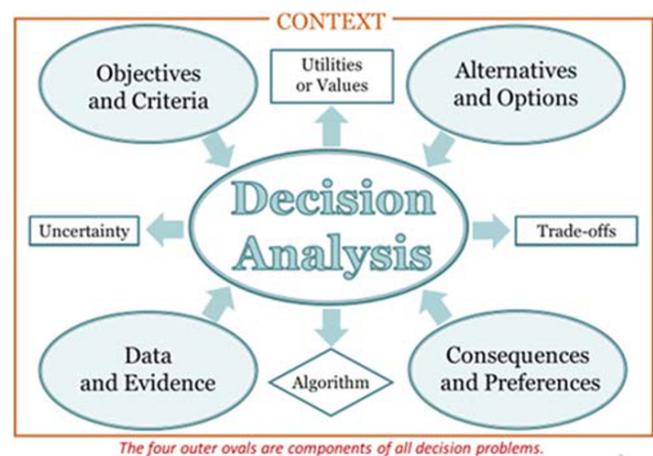

FIGURE 2 Qualitative, semi-quantitative and fully quantitative decision analysis approaches to benefit–risk assessment. Framework 1: Four outer ovals = qualitative approach. Framework 2: + three top rectangles = semi-quantitative approach. Framework 3: + single bottom diamond = fully-quantitative approach

analyses in a fully quantitative model can often provide agreement about the decision without requiring consensus about the details. As for the resources required for a fully-quantitative multi-criteria model, EMA feasibility studies on five drugs that were being considered for approval by the CHMP in 2010–2011, were conducted with the participation of four to six experts (including the rapporteur or co-rapporteur) within a period of six hours[21] (more information on these initiatives is provided below).

Some might wonder whether the fully quantitative approach using an algorithm is necessary, or whether the aggregation step could be left to the decision-makers. Evidence to answer these questions can be found in the cognitive psychology literature. As famously pointed out by psychologist George Miller over half a century ago, the human brain can keep in mind at one time about five to nine pieces of information.[22] Two years earlier, Paul Meehl showed that simple, linear, additive models consistently outperformed clinical predictions of patient behaviour.[23] Although originally based on a sample of 20 studies, by 1996 an analysis of 136 comparative studies came to the same conclusion, with only a small proportion of the studies (around 5%) favouring clinical prediction, the underlying problem identified as the limited human "integration" of multiple pieces of evidence. More recent scholars in behavioural economics like Thaler and Sunstein questioned the rationality of human judgements,[24] together with Kahneman who acknowledged that the human brain lacks such an "integrator" or that the human integrator has limited capacity.[25] Limited mental capacity causes an assessor to focus on a single or small number of effects, which Montibeller and von Winterfeldt[26] label as 'myopic problem representation'. In addition, risk attitude and other biases were exhibited by 80 European medical assessors answering online questionnaires about hypothetical investigational drugs treating central nervous system, cardiovascular and oncology disorders.[27] These biases can be minimized by applying group elicitation techniques for quantifying the subjective components.[26]

In multi-criteria decision analysis (MCDA), "algorithms" simply aggregate the components to give an overall benefit–safety balance, i.e. 'handing back' the inputs in changed form, which often stimulates new insights about the benefit–safety balance. Furthermore, as the model is explored to see why the overall result was obtained, the process helps participants to construct their value preferences. All this is more easily accomplished with numbers than with words alone, and it provides the group with sufficient clarity for an informed decision to be taken. The model only aggregates; it does not replace human judgement, nor does it dictate the 'right' solution.

MCDA is a methodology for weighing options on individual, often conflicting criteria, and combining them into one overall appraisal, taking into account both the value of outcomes and their uncertainty, as well as all the items in Figure 2. Howard Raiffa, the founder of decision analysis, described the spirit of the process as "divide and conquer: decompose a complex problem into simpler problems, get one's thinking straight on these simpler problems, paste these analyses together with logical glue, and come out with a program of action for the complex problem."[28] The logical glue is the algorithm for combining the components of the model.

In the 1960s and 1970s, decision analysis texts and applications focused on problems involving uncertainty,[29] but extensions to decision theory broadened the discipline to include decisions with multiple, often conflicting, objectives.[30] MCDA now encompasses both uncertainty and value. Using MCDA to model the benefit–risk balance of drugs would include representing uncertainties as probabilities and values defined as numerical measures of the extent to which outcomes realize their associated objectives. The algorithm for combining these two is to multiply value by probability, just as you would judge the fair value of a 50–50 chance of winning $1,000 to be $500.

Since algorithms usually outperform unaided human judgement, and explicit quantification can reduce bias among assessors, in the next section, we will describe how such algorithms can be employed in regulatory decision-making.

4 | EUROPEAN EXPERIENCE WITH QUANTITATIVE BENEFIT–RISK MODELLING AND THE PROACT-URL FRAMEWORK

Two extensive reviews, conducted under the auspices of the EMA (as part of the Benefit–Risk Methodology Project and the IMI-PROTECT project), examined the suitability of a variety of fully quantitative models for regulatory decision-making.[7,31] The EMA's CHMP has explored the use of MCDA for modelling drug benefit–risk balance. The PROTECT project concluded that "All teams chose MCDA … because of its comprehensiveness, accommodation of any effect metrics and value judgements, and support for trade-off weighting, all requirements for a fully quantitative model."[7] Although numerous MCDA approaches and methods exist, they share most of their key steps.[32] Perhaps the most generic framework is PrOACT-URL,[33] which was adapted for benefit–risk assessment as part of EMA's Benefit Risk Project.[21] This framework, outlined in Table 1, acted as the progenitor of a series of qualitative frameworks that emerged for the evaluation of drug benefit–risk balance,[34–36] including FDA's own BRF recommendations.[4]

Benefit–risk MCDA models do not need to be very complex. The concept of requisite decision modelling can be adopted, defined as the model for which the form and content are sufficient to solve a particular problem.[37] This can be achieved through a 'socio-technical decision analysis' approach, in which the technical decision-analytic modelling and the group discussion are in a reflexive relationship, each supporting the other, leading to a shared understanding among participants of the issues that inform the final decision.[38] For example, a preliminary complex model can be trimmed down by eliminating effects that negligibly affect the overall benefit–risk balance, identified by sensitivity analyses that explore imprecision of the data and differences in judgements. Ideally, an impartial facilitator guides the process as part of face-to-face workshops[39] or decision conferences.[40] Current best practices for MCDA in health care, including the appropriate identification and selection of criteria, are extensively discussed elsewhere.[41,42]

TABLE 1 The PrOACT-URL framework for drug benefit–risk assessment

Problem	Determine the nature of the problem and its context.
	Frame the problem.
Objective	Establish objectives that indicate the overall purposes to be achieved.
	Identify criteria for (a) favourable effects, and (b) unfavourable effects.
Alternatives	Identify the options to be evaluated against the criteria.
Consequences	Describe how the alternatives perform for each of the criteria, i.e., the magnitudes of all effects, their desirability or severity, and the incidence of all effects.
Trade-offs	Assess and report the balance between favourable and unfavourable effects.
Uncertainty	Report the uncertainty associated with the favourable and unfavourable effects.
	Consider how the balance between favourable and unfavourable effects is affected by uncertainty.
Risk tolerance	Judge the relative importance of the decision-maker's risk attitude for this product.
	Report how this affected the balance reported above.
Linked decision	Consider the consistency of this decision with similar past decisions and assess whether taking this decision could impact future decisions.

This facilitated-group process has been applied by several pharmaceutical manufacturers to compare their products with other drugs for the same medical condition. Recent examples include Reckitt Benckiser for over-the-counter analgesics,[43] Merck Serono for multiple sclerosis[44] and Pfizer for post-operative analgesics and overactive bladder.[46] In every case, the company explored the literature to establish three to five favourable effects and up to 11 or 12 unfavourable effects, gathered the associated data, and created a model in a decision conference of company and external experts and clinicians, facilitated by one of the co-authors. The independent external experts assessed trade-off values, made changes and provided an impartial validation of the model and its results. In every case, the overall results revealed features about the differences among the drugs that had not been evident from the individual data, which would enable prescribers to better target a drug for an individual patient. For example, in the overactive bladder case, the benefit–risk balances of two drugs were shown in the MCDA to be worse than the placebo. The high safety score for the placebo plus its small benefit exceeded the sums of the small-to-modest benefits and poor safety of the drugs.

Among the main benefits of MCDA in the evaluation of new medicines is its comprehensiveness in terms of enabling simultaneous incorporation of several dimensions of value of the favourable effects and loss of value by the unfavourable effects. Additionally, it is possible to test the effects of stakeholders' views about clinical relevance, including patients.[46] Importantly, benefits include the ability to facilitate the constructive thinking of decision-makers about their own preferences[47] and value trade-offs between different evaluation criteria, therefore improving transparency and consistency of decision outcomes.

A number of misinterpretations could be raised that are perceived to act as challenges for MCDA applications in benefit–risk assessment, relating to limitations of clinical trial evidence. For example, the lack of a randomized control arm in a drug's pivotal clinical study would not be something for MCDA to solve. Clearly, any analysis involving some type of modelling is as good as the input data used to populate the model, however various MCDA features could be used to mitigate the impact of such limitations. For example, sensitivity analyses of confidence interval limits could be sufficient to address variability among individual patient-level effects, whereas probabilistic sensitivity analysis could be used to identify potential safety issues to be included in risk mitigation programs. Past experience in the field indicates that experts participating in MCDA studies are often happy to handle evidence limitation and data quality issues by assigning less weight to criteria for which data are considered to be poor.[41]

Interestingly enough, MCDA and decision conferencing had been acknowledged by the US National Academies of Science, Engineering and Medicine's Committee on Ethical and Scientific Issues in Studying the Safety of Approved Drugs since 2012 as the only methodology that could lead to a consistent decision-making framework across the lifecycle of new drugs, while also allowing for input from patients and other stakeholders (Appendix, Recommendation 2.1).[48]

5 | EMA CASE STUDY: BELIMUMAB FOR SYSTEMIC LUPUS ERYTHEMATOSUS

The following section illustrates an application of the PrOACT-URL process in developing a fully quantitative decision-analytic model with EMA regulators and clinical experts. More precisely, this case study focuses on belimumab (Benlysta), when it was first reviewed for approval by the EMA for the indication of systemic lupus erythematosus (SLE). We believe this is the first case to report how a fully quantitative model helped to resolve differences of judgements among drug regulators. It has been disguised but faithfully reports the process and results as reported in publicly available documents.[49,50] Interested readers could seek further information on MCDA for medicinal products and health care decisions elsewhere.[38,41]

1 | Problem

Early in 2011, EMA regulators preparing the 150-day report were finding it difficult to agree about whether belimumab (Benlysta), the first new drug in 56 years for treating SLE, should be approved. All agreed the drug's benefits were modest but disagreed about some safety issues.

2 | Objective

A one-day decision conference was convened to determine if a decision analysis model of the benefit–risk balance of belimumab could help to resolve the disagreements among regulators sufficiently for them to agree a recommendation to the CHMP. Two clinical assessors, one non-clinical assessor, a quality assessor and a pharmacist, all of whom were familiar with the data about belimumab, served as experts in the decision conference, which was facilitated by one of the co-authors and assisted by an EMA assessor who provided computer modelling support.

3 | Alternatives

1. Belimumab 10 mg
2. Belimumab 1 mg
3. Placebo

4 | Consequences

The group agreed the six favourable and three unfavourable effects, i.e. evaluation criteria, shown in the drug's Effects Table (Table 2); the effects' operational definitions, their measurement units and the performance of the alternative treatment options are also shown. For the selection of the effects, the decision conference facilitator asked questions to ensure the effects were preference independent, even in cases of a common cause. To provide a common metric for all effects, all data were converted to 0–100 preference value scales, with zero assigned to the least well-performing drug and 100 to the best-performing drug on each scale. Conversions were either direct linear transformations (larger numbers are more preferred) or indirect linear (smaller numbers are more preferred, as for all unfavourable effects). An exception to the linear transformation was Flare rate, for which a non-linear value function was deemed more appropriate over the whole range from 0 to 5 cases per patient year, as shown in Figure3. The value function shows that the loss of value from no flares is 10 points for one flare, 30 for two, 40 for three and levelling off thereafter.
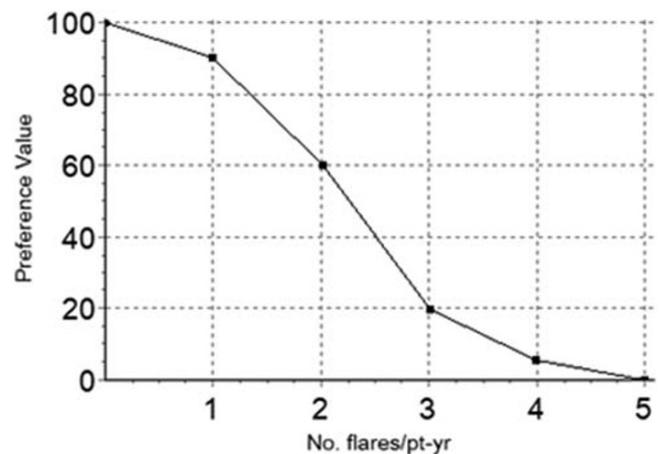
FIGURE 3 Value function of the flare rate criterion

TABLE 2 Effects, definition, units and performance used in the MCDA case study with EMA regulators on the use of belimumab for systemic lupus erythematosus

Effect	Definitions	Units	Weight	10 mg	1 mg	PBO
Favourable effects						
SLEDAI 4–6% improved	Percentage of patients with at least 4 points reduction in SLEDAI at week 52	%	2.3	16	15	18
SLEDAI > 6% improved	Percentage of patients with more than 6 points reduction in SLEDAI at week 52	%	5.9	37	33	23
PGA % no worse	Percentage of patients with no worsening in Physician's Global Assessment	%	1.2	75	76	66
BILAG A/B	Percentage of patients with no new BILAG A/B	%	3.5	75.2	70.1	69.0
CS sparing	Percentage of patients that reduced the dose of corticosteroids (CS) by more than 25% and to less than 7.5 mg/day	%	3.9	15.5	20.0	12.3
Flare rate	Number of new BILAG A cases per patient-year	Number	39.1	2.88	2.90	3.51
Unfavourable effects						
Potential serious adverse events (SAEs)	Potential for developing tumour, adverse interactions with vaccines and AE on pregnancies	Direct judgement	16.0	0	90	100
Infections	Proportion of patients with serious infections that are life-threatening	%	23.4	5.2	6.8	5.2
Sensitivity reaction	Proportion of patients with hypersensitivity reactions	%	4.7	0.4	1.3	0.1

SLEDAI: Systemic Lupus Erythematosus Disease Activity Index; PGA: Physician's Global Assessment; BILAG: British Isles Lupus Assessment Group. All scales are defined in Mosca, M., Bombardieri, S. Assessing remission in systemic lupus erythematosus. Clin Exp Rheumat. 2006;24(Suppl. 43):S100-S104.

5 | Trade-offs

The purpose of weighting in decision theory is to ensure that the units of preference value on the different scales are equivalent, thus enabling weighted scores to be compared and combined across the criteria. Weights are scale factors that represent the extent of trade-offs between criteria, as 9 Fahrenheit units equate to 5 Celsius units of temperature.

To assess criteria weights, the process of swing-weighting was used. Two steps in thinking must be separated. First, it is necessary to consider the objective difference in effect size between the least and most preferred effects on a given criterion; that is available from the evidence. The next step is to think about how much that difference matters; this is essentially a judgement of the clinical relevance of the difference in effect size. "How big is the difference and how much do you care about that difference?" This is the question that was posed in comparing the 0-to-100 swing in effect on one scale with the 0-to-100 swing on another scale, usually comparing only two effects at a time.

During the assessment process, the facilitator applied various techniques to minimize bias in making these necessarily subjective judgements. For example, ratios of weights are compared: the weight on the primary endpoint, SLEDAI, is the sum of the weights on its two criteria, 8.2, and that was judged to be about twice as much added benefit as for CS sparing at 3.9. Flare rate's weight of 39.1 is a little more than twice SAE's 16. Participants usually revised their original assessments after these checks. The final weights are shown in the Effects Table (Table 2). Note that a weight is not the importance of the criterion; instead, it represents the added clinical value from least preferred to most preferred in the context of the other criteria.

With preference values established and weights agreed, the computer simply multiplies the values by the weights (after dividing them by 100) and those weighted values are summed, representing a linear additive model. The resulting weighted preference values are interpreted as benefits for the favourable effects and safety for the unfavourable effects.

Figure 4 shows overall weighted preference value scores, with favourable effects, i.e. benefits, as green and unfavourable effects,

i.e. safety, as red (so that more red means safer). The 1 mg drug shows an overall positive balance, with too little benefit associated with the placebo and too little safety for the 10 mg dose. Clinical judgement captured in the model favoured the 1 mg dose as a reasonable compromise between the stronger dose and the placebo. However, the overall benefit–safety balance scores are very close to each other, reflecting the difficulty experts were experiencing in agreeing about belimumab. Figure 5 shows the contribution of each criterion to the overall scores. It is clear that the largest contributors to safety are potential SAEs and infections, with some differences in sensitivity reactions.

6 | Uncertainty

Weights are necessarily subjective judgements and participants felt uncertain about the numbers they had agreed. A sensitivity analysis shows how increasing or decreasing the weight on an effect results in a change in the overall benefit–safety balance. Sensitivity analyses on all the unfavourable effects left assessors with a dilemma: a slight increase in the weight on sensitivity reaction, as shown in Figure 6, favours the placebo, or a slight decrease in weight for potential SAEs, as shown in Figure 7, favours the 10 mg dose. Clinical judgement captured in this model favoured the 1 mg dose as a reasonable compromise between the stronger dose and the placebo, and that might have resolved the experts' disagreements.

However, no SAEs were actually observed in the clinical trials. By setting the weight on potential SAEs to zero, the model showed overall preference values of 46, 33 and 33 for Benylsta 10 mg, 1 mg and placebo, respectively, a clear win for the 10 mg dose. The model helped to resolve disagreements among the regulators and experts, and three months later, the CHMP gave a positive opinion about belimumab (though they had not been shown the model). More precisely, some regulators had argued for the 10 mg dose based on its superior benefits over the 1 mg dose, whereas others felt that the potential for SAEs in the 10 mg dose could lead to an unsafe approval. By removing SAEs from consideration, the 10 mg dose showed a

FIGURE 4 Overall weighted preference value scores of treatment options, illustrating benefits (green) vs. safety (red)
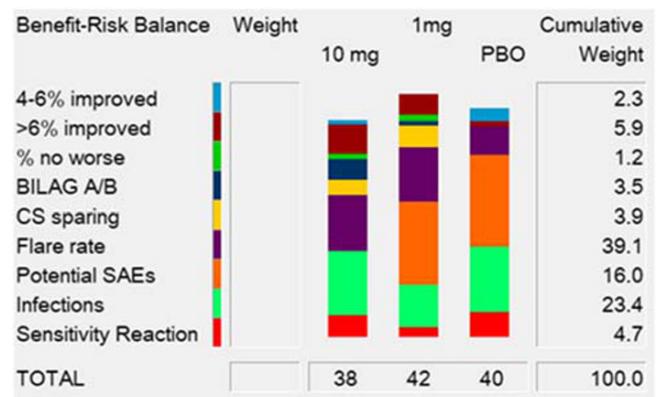

FIGURE 5 Overall weighted preference value scores of treatment options, illustrating individual criteria contributions (different colours)
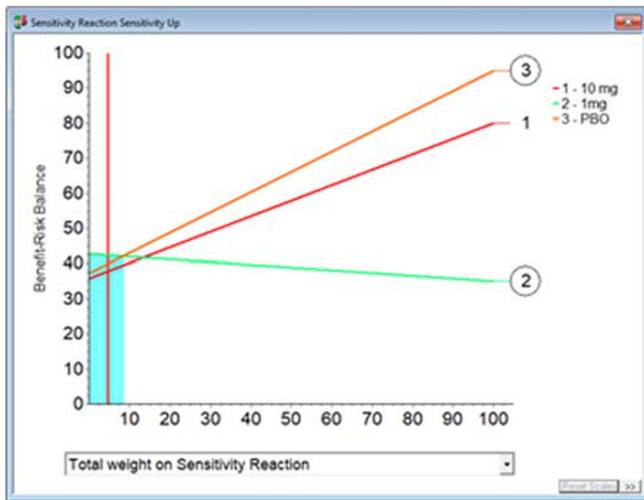
FIGURE 6 Sensitivity analysis on the weight of the sensitivity reaction criterion.
The vertical red line is located at the current weight for that effect (4.7) and it intersects the sloping lines at their current overall preference values (38 for 10 mg, 42 for 1 mg, 40 for PBO)

6-point lead over the 1 mg dose, and regulators agreed that the level of benefit sufficiently outweighed the risk, especially as no SAEs had been observed in the clinical studies. This is a good example of how numbers can deepen understanding of disagreements that are not easily resolvable with words.
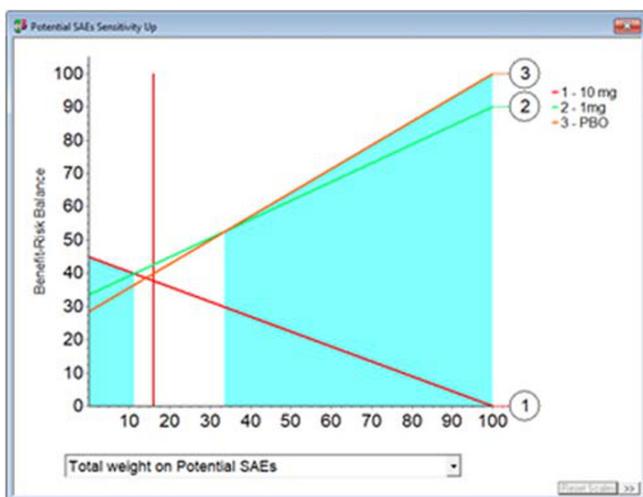

FIGURE 7 Sensitivity analysis on the weight of the potential SAEs criterion.
The vertical red line is located at the current weight for that effect (16.0) and it intersects the sloping lines at their current overall preference values (38 for 10 mg, 42 for 1 mg, 40 for PBO)

7 | Risk tolerance

The analysis revealed the importance of possible serious adverse events and how these could substantially affect the benefit–safety balance. The European Public Assessment Report emphasized the importance of a long-term follow-up study.

8 | Linked decision

There is no link to past decisions, but in future this analysis could serve as a starting point for modelling new drugs about lupus, as for belimumab the decisions about what effects to consider were based solely on existing drugs to treat the medical condition (which is why the potential for SAEs was included) as well as the clinical findings.

One issue raised in the decision conference was the double counting of patients in the two SLEDAI scores, which were originally defined as 'At least a 4-point reduction' and 'More than 6-points reduction'. The percentage of patients satisfying the latter also satisfies the former, so some patients were counted twice, exaggerating the impact of the reduction. Here the definitions that appear in the above Effects Table result in a simple frequency distribution, and the weight on SLEDAI 4–6% was reduced to accommodate the range of smaller percentage reductions.

6 | MCDA FOR REGULATORY BENEFIT–RISK RECOMMENDATIONS

The main feature of MCDA is that it provides a way for transforming the objective measure of a drug's performance into a common metric of preference value across all its effects versus one or more comparators. This is accomplished by establishing plausible ranges of the effects data and making informed judgements about the clinical relevance of those ranges. With all effects transformed to a common metric of preference value, and weights assigned about the clinical relevance of the ranges of data, it becomes possible to add benefits to safety, thus providing an overall index number of added clinical value for a drug.

This explicit process and its results improve the "clarity and consistency in communicating the reasoning behind drug regulatory decisions" that are the objectives of the BRF in the following ways: (i) listing the favourable and unfavourable effects makes clear which benefits and risk criteria were considered by the regulator, (ii) the list of effects provides guidance to pharmaceutical companies of what matters (and, perhaps more importantly, what does not matter) to the regulators, and it ensures consistency for the regulator in dealing with new drugs for the same medical condition, (iii) weights make explicit the extent of risk attitudes by the assessors, which helps to reduce bias in the model, and (iv) MCDA uses an algorithm for combining the pieces making up the benefit–safety balance, which overcomes the inherent bias of focusing on a single or small number of effects.

The European experience with EMA on the use of quantitative decision analysis following the Benefit–Risk Methodology Project and the IMI-PROTECT projects has demonstrated the prospects of using these methods for assessing drug benefit–risk balance and facilitating regulatory decisions. Today, although the EMA does not conduct or place a requirement for quantitative decision analytic methods, it accepts the submission of such evidence and conducts the review of relevant studies. More tangibly though, the EMA encourages the use of quantitative frameworks via the full implementation of the Effects Table, which forms a fundamental task in data collection, synthesis and analysis of evidence as part of quantitative decision analysis methods.

Beyond that, it would be challenging to identify the impact that any such studies might have had so far on informing regulatory decisions: EMA regulators do not report in the European Public Assessment Reports (EPARs) how they come to make their decisions; CHMP meetings are not transcribed and are not open to the public; and EMA guidance documents explain what must be taken into account and reported, but not how the information will be put together. The disclosure of value judgements and value preferences seems to be a very sensitive issue.

Further elaboration of the FDA's BRF elements could benefit from the eight-step PrOACT-URL process following the European experience, either by adopting it in full, in part or just by using a similar approach. For example, in terms of partial adoption, incorporating an ET into public documents after a new drug is evaluated would improve the transparency and communicability of the regulatory decision, as it has for the EMA. Indeed, the discipline of constructing the table has proved to clarify the thinking of the regulators. In the case of full adoption, regulatory communication could be further improved by the systematic disclosure of the model used, such as the value tree of favourable and unfavourable effects that supported the decision. Whilst judgements and opinions may vary over time and geographical regions, this would create a major opportunity for a transparent dialogue and appraisal of clinical relevance.

Besides assessing drug benefit–risk balance for regulatory approval, quantitative decision analysis could also be used for other regulatory decisions such as the appropriate timing of a vaccine's approval for a pandemic. An example would be a quantitative modelling application on the risk–benefit impact of H1N1 influenza vaccines, to inform the decision between approval based on limited data or waiting for more data to become available.[51] The findings suggested that quantitative models might be helpful to regulators for such public health issues characterized by considerable uncertainty, thus making them directly relevant to the current Covid-19 pandemic situation.

If a quantitative model is to be constructed, who should do it? An interesting parallel is provided by what the FDA and EMA require for a new product's submission: statistical data related to the primary endpoint and, possibly, of secondary endpoints. Each organization's statisticians are available to comment on experimental design and statistical methods used to analyse the data. Perhaps the same capability should exist for MCDA modelling, with each applicant supplying its MCDA model and the regulator critiquing the submitted MCDA model if not constructing its own model from the submitted data. In either case, the regulator could then explore differences of opinion, value judgements, and trade-off assessments by conducting sensitivity analyses. Ultimately, agencies will need to become competent at interpreting MCDA models for their validity, as they have learned to identify good statistical practice.

7 | CONCLUSION

While the FDA has recently started to implement a qualitative decision framework, the EMA has been testing quantitative frameworks for about a decade and now will take note of any quantitative model of the benefit–risk balance in regulatory submissions of applications for marketing authorization of medicinal products. If the FDA decides to move towards a more quantitative framework based on the decision sciences literature and the past European experience of benefit–risk methodologies in drug regulatory decision-making, it is apparent that structured decision analysis approaches such as MCDA could accommodate all necessary features of a solid and robust benefit–risk assessment. Other ongoing FDA initiatives relating to incorporation of patient preferences, the MIDD Pilot Program, and RWE, would act as further opportunities for the use of a quantitative framework as it could consolidate their outputs and enhance drug regulatory decision-making.

ORCID
Aris Angelis https://orcid.org/0000-0002-0261-4634

REFERENCES
1. EMA. Benefit-Risk methodology. European Medicines Agency. http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000314.jsp&mid=WC0b01ac0580981014. Published 2019. Accessed May 2019.
2. Califf RM. Benefit-risk assessments at the US Food and Drug Administration: finding the balance. *JAMA*. 2017;317(7):693-694.
3. FDA. Benefit-Risk Assessment in Drug Regulatory Decision-Making, Draft PDUFA VI Implementation Plan (FY 2018–2022). 2018.

4. Jenkins J. A United States regulator's perspective on risk-benefit considerations. Risk-benefit considerations in drug regulatory decision-making. Paper presented at Shady Grove Conference Center 2010; Rockville, MD

5. FDA. Statement from FDA Commissioner Scott Gottlieb, M.D., on FDA's efforts to enhance the patient perspective and experience in drug development and review. 2018.

6. Eichler HG, Abadie E, Raine JM, Salmonson T. Safe drugs and the cost of good intentions. *N Engl J Med*. 2009;360(14):1378-1380.

7. Hughes D, Waddingham E, Mt-Isa S, et al. Recommendations for benefit–risk assessment methodologies and visual representations. *Pharmacoepidemiol Drug Saf*. 2016;25(3):251-262.

8. Edwards W, Lindman H, Phillips LD. Emerging technologies for making decisions. *New Directions in Psychology Volume 2*. New York, NY: Holt, Rinehart and Winston; 1965.261.

9. Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern Med*. 2013;173(8):611-612.

10. EMA. Advanced therapy medicinal products: overview. https://www.ema.europa.eu/en/human-regulatory/overview/advanced-therapy-medicinal-products-overview; 2020.

11. Kesselheim AS, Avorn J. Approving a problematic muscular dystrophy drug: implications for FDA policy. *JAMA*. 2016;316(22):2357-2358.

12. Darrow JJ. Luxturna: FDA documents reveal the value of a costly gene therapy. *Drug Discov Today*. 2019;24(4):949-954.

13. Carpenter D, Kesselheim AS, Joffe S. Reputation and precedent in the bevacizumab decision. *N Engl J Med*. 2011;365(2):e3(1)–e3(3).

14. FDA. Model-Informed Drug Development Pilot Program. 2020. https://www.fda.gov/drugs/development-resources/model-informed-drug-development-pilot-program

15. FDA. Real-World Evidence. 2020. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

16. Phillips L, Fasolo B, Zafiropoulos N, Beyer A. Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug Discov Today Technol*. 2011;8(1):e1-e42.

17. EMA. <Co>Rapporteur day <60*><80> critical assessment report. 2018.

18. EMA. <Co->Rappoeteur day 80 critical assessment report. (EMA/90842/2015). 2015.

19. FDA. Agenda Item: FDA Risk Analysis of the OraQuick In-Home HIV Test Based on Clinical Trial Performances, available in FDA Center for Biologics Evaluations and Research. Paper presented at 102nd Meeting of The Blood Product Advisory Committee (BPAC) 2012.

20. Morales JP, Li X, Irony TZ, Ibrahim NG, Moynahan M, Cavanaugh KJ. Decision analysis of retrievable inferior vena cava filters in patients without pulmonary embolism. *J Vasc Surg Venous Lymphat Disord*. 2013;1(4):376-384.

21. EMA. Benefit-risk methodology project. Work package 3 report: Field tests. 2011; EMA/227124/2011:29. Accessed 26 July 2016.

22. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev*. 1956;63(2):81-97.

23. Meehl PE. *Clinical versus statistical prediction: a theoretical analysis and a review of the evidence*. Minneapolis: MN: University of Minnesota Press; 1954.

24. Thaler RH. *Nudge: improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press; 2008.

25. Kahneman D. *Thinking, fast and slow*. London: Allen Lane; 2011.

26. Montibeller G, Winterfeldt D. Cognitive and motivational biases in decision and risk analysis. *Risk Anal*. 2015;35(7):1230-1251.

27. Beyer AR, Fasolo B, de Graeff PA, Hillege HL. Risk attitudes and personality traits predict perceptions of benefits and risks for medicinal products: a field study of European medical assessors. *Value Health*. 2015;18(1):91-99.

28. Raiffa H. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley; 1968: 271.

29. Decision Analysis Group. *Readings in Decision Analysis*. 2nd ed. Stanford, CA: Stanford Research International; 1976.

30. Keeney R, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Cambridge: Cambridge University Press; 1993.

31. EMA. *Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment*. London: European Medicines Agency; 2010:33.

32. Dodgson J, Spackman M, Pearman A, Phillips L. *Multi-Criteria Analysis: A Manual*. London: Department for Communities and Local Government; 2009:2000.

33. Hammond J, Keeney R, Raiffa H. *Smart Choices: A Practical Guide to Making Better Decisions*. Cambridge, MA: Harvard University Press; 1999.

34. CIRS. The CIRS-BRAT Framework, Centre for Innovation in Regulatory Science. 2019.

35. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin Pharmacol Ther*. 2009;85(3):241-246.

36. Liberti L, McAuslane N, Walker S. Progress on the development of a benefit/risk framework for evaluating medicines. *Reg Focus*. 2010;15:32-37.

37. Phillips L. A theory of requisite decision models. *Acta Psychol (Amst)*. 1984;56(1-3):29-48.

38. Phillips LD. Benefit-risk modeling of medicinal products: Methods and applications. In: Sashegyi A, Felli J, Noel R, eds. *Benefit-Risk Assessment in Pharmaceutical Research and Development*. Boca Raton, FL: CRC Press; 2014.

39. Franco LA, Montibeller G. Facilitated modelling in operational research. *Eur J Op Res*. 2010;205(3):489-500.

40. Phillips L. Decision conferencing. In: Edwards W, Miles R, von Winterfeldt D, eds. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge: Cambridge University Press; 2007.

41. Phillips LD. Best practice for MCDA in healthcare. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, eds. *Multi-Criteria Decision Analysis to Support Healthcare Decisions*. New York, NY: Springer; 2017:311-329.

42. Marsh K, Ijzerman M, Thokala P, et al. Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ISPOR MCDA Emerging Good pPractices Task Force. *Value in Health*. 2016;19(2):125-137.

43. Moore A, Crossley A, Ng B, Phillips LD, Sancak Ö, Rainsford KD. Use of multicriteria decision analysis for assessing the benefit and risk of over-the-counter analgesics. *J Pharm Pharmacol*. 2017;69(10):1364-1373.

44. Vermersch P, Martinelli V, Pfleger C, et al. Benefit-risk assessment of cladribine using multi-criteria decision analysis (MCDA) for patients with relapsing-remitting multiple sclerosis. *Clin Ther*. 2019;41(2):249-260.

45. Schug S, Pogatzki-Zahn E, Phillips LD., et al. Multi-criteria decision analysis to develop an efficacy-safety profile of parenteral analgesics used in the treatment of postoperative pain. *Journal of Pain Research*. 2020; in press.

46. ChappleChristopher R., Mironska Emma, Wagg Adrian, Milsom Ian, Diaz David Castro, Koelbl Heinz, Pushkar Dmitry, Tubaro Andrea, De Ridder Dirk, Chartier-Kastler Emmanuel, Phillips Lawrence D.. Multi-criteria Decision Analysis Applied to the Clinical Use of Pharmacotherapy for Overactive Bladder Symptom Complex. *European Urology Focus*. 2020;6 (3):522–530. https://doi.org/10.1016/j.euf.2019.09.020.

47. Angelis A, Kanavos P. Value-based assessment of new medical technologies: towards a robust methodological framework for the application of multiple criteria decision analysis in the context of health technology assessment. *Pharmacoeconomics*. 2016;34(5):435-446.

48. Lichtenstein S, Slovic P. The Construction of Preference: An Overview. *The construction of preference*. Cambridge: Cambridge University Press; 2006.1–40.

49. Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*. Washington, DC: The National Academies Press; 2012.

50. European Medicines Agency. Work package 4 report: Benefit-risk tools and processes. London: European Medicines Agency; 2012.

51. European Medicines Agency. *Assessment report Benylsta (belimumab)*. London: European Medicines Agency; 2015 WC500110152 EPAR.

52. Phillips LD, Fasolo B, Zafiropoulous N, et al. Modelling the risk–benefit impact of H1N1 influenza vaccines. *Eur J Public Health*. 2013; 23(4):674-678.