Drug Saf (2015) 38:1049–1057
DOI 10.1007/s40264-015-0344-2

CURRENT OPINION

A Note on the Validity and Reliability of Multi-Criteria Decision Analysis for the Benefit–Risk Assessment of Medicines

Alberto Garcia-Hernandez[1]

Published online: 9 September 2015

Abstract The comparative evaluation of benefits and risks is one of the most important tasks during the development, market authorization and post-approval pharmacovigilance of medicinal products. Multi-criteria decision analysis (MCDA) has been recommended to support decision making in the benefit–risk assessment (BRA) of medicines. This paper identifies challenges associated with bias or variability that practitioners may encounter in this field and presents solutions to overcome them. The inclusion of overlapping or preference-complementary criteria, which are frequent violations to the assumptions of this model, should be avoided. For each criterion, a value function translates the original outcomes into preference-related scores. Applying non-linear value functions to criteria defined as the risk of suffering a certain event during the study introduces specific risk behaviours in this prescriptive, rather than descriptive, model and is therefore a questionable practice. MCDA uses weights to compare the importance of the model criteria with each other; during their elicitation a frequent situation where (generally favourable) mild effects are directly traded off against low probabilities of suffering (generally unfavourable) severe effects during the study is known to lead to biased and variable weights and ought to be prevented. The way the outcomes are framed during the elicitation process, positively versus negatively for instance, may also lead to differences in the preference weights, warranting an appropriate justification during each implementation. Finally, extending the weighted-sum MCDA model into a fully inferential tool through a probabilistic sensitivity analysis is desirable. However, this task is troublesome and should not ignore that clinical trial endpoints generally are positively correlated.

Key Points

Both the European Medicines Agency and the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium have proposed multi-criteria decision analysis (MCDA) for drugs-related benefit–risk assessment (BRA).

The weighted-sum MCDA model is exposed to a number of potential sources of bias or variability in this area.

MCDA can be a valid and reliable tool for decision making in BRA provided its implementation adheres to the assumptions of the model with scientific rigour.

1 Introduction

Although determining the benefit–risk balance of a new medicine is one of the most important tasks during its development, market authorization and post-approval reassessment, it is only recently that regulatory agencies, pharmaceutical companies and other groups have begun to discuss and pilot methods to establish transparent processes and standardize the benefit–risk assessment (BRA) of

✉ Alberto Garcia-Hernandez
alberto.garciahernandez@astellas.com

[1] Global Data Science, Astellas Pharma Europe B.V., Sylviusweg 62, 2333 BE Leiden, The Netherlands

medicines. The risk–benefit management working group of the International Society of Pharmacoeconomics and Outcome Research (ISPOR), the European Medicines Agency (EMA) benefit–risk methodology project (BRMP) and the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) have published three reviews of quantitative methods that could potentially be used for a drug's BRA [1–3]. With regard to the assessment performed for marketing authorisation, the EMA BRMP has recommended multi-criteria decision analysis (MCDA) for difficult or contentious cases [4], whereas the US FDA appears to have a different view, warning that reliance on a relatively complex model might obscure rather than elucidate a regulator's thinking and thus has favoured a qualitative approach [5]. In the field of pharmacoepidemiology and pharmacovigilance, the PROTECT programme on BRA identified up to 49 methodologies that could be used to balance the benefits and risks of drugs and selected both MCDA and stochastic multicriteria acceptability analysis (SMAA) for further examination [3, 6].

MCDA was initially proposed for the BRA of medicines by Mussen et al. [7, 8, 9] and has been piloted for the evaluation of several compounds [10, 11].

This paper is structured as follows. Section 2 provides an introduction to weighted-sum MCDA for BRA, illustrated with an implementation in rheumatoid arthritis. Section 3 reviews several challenges faced by MCDA in this area and provides solutions to overcome them; and, finally, Sect. 4 concludes the paper. This work does not aim to be an exhaustive inventory of good practices using MCDA but represents only the author's views on some key points that, if not adequately addressed, jeopardize the validity and reliability of MCDA in the BRA field.

2 The Weighted-Sum Multi-Criteria Decision Analysis (MCDA) Model for Benefit–Risk Assessment (BRA)

The weighted-sum MCDA model evaluated in this paper belongs to a group known as value measurement models [12] and constructs an overall value V_j for each compound j as follows:

$$V_{\mathrm{j}} = \sum_{\mathrm{i}=1}^{\mathrm{m}} w_{\mathrm{i}} v_{\mathrm{i}}(x_{\mathrm{ij}}) \qquad (1)$$

where i is indexed over m drug criteria, w_i is the importance weight of the ith criterion, x_{ij} is the outcome on criterion i with drug j and $v_i(\cdot)$ is the score or partial value function on that criterion. One important feature of the way MCDA has been proposed for BRA is that criteria are defined in terms of group-level responses, in other words, the model is constructed on the basis of efficacy and safety summaries from the study reports rather than on subject-level outcomes. This approach contrasts with other decision tools such as the quality-adjusted life-year (QALY) model, where preferences are assigned at an individual health-outcome level. This framework is particularly flexible, since a group of stakeholders can take the available results, which may have been analyzed and reported using heterogeneous methodologies, from one or more clinical trials and, without the need to use the subject-level database or complex modelling, fully specify a decision model that supports their assessment.

2.1 Value Tree and Identification of Criteria

The initial task in any decision process is framing the problem context. In BRA, this refers to the identification of elements such as indication, medical need, target population and available therapeutic options. Once the decision problem has been contextualized, we need to identify a set of performance criteria associated with favourable and unfavourable effects that should satisfy certain characteristics such as relevance, completeness, non-redundancy, understandability, feasibility, mutual independence of preferences and non-overlap [7, 12]. Figure 1 depicts the value tree defined for an MCDA model in rheumatoid arthritis implemented by the EMA BRMP [10, 13]. Table 1 shows how criteria were defined in terms of the clinical trial summaries, and a range of possible consequences was used as reference to quantify the preferences over the responses of each criterion. The range of possible outcomes for the primary efficacy endpoint, American College of Rheumatology (ACR)-20 response [14], was defined as

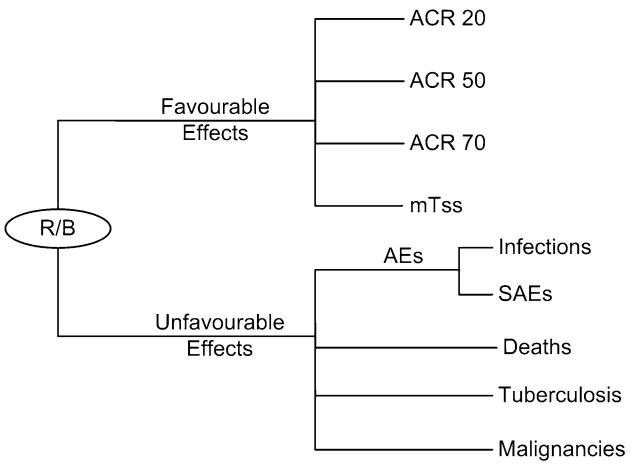

Fig. 1 Value tree in rheumatoid arthritis [10]. Reproduced and adapted from the benefit-risk methodology project work package 3 report: field tests [10]. *ACR* American College of Rheumatology, *AEs* adverse events, *mTSS* modified total Sharp score, *R/B* risk/benefit, *SAEs* serious adverse events

Table 1 Multi-criteria decision analysis model in rheumatoid arthritis [10]

Criteria	Definition (units)	Range of outcomes	Outcome (value)			Weight
			Placebo	Active (200 mg)	Active (400 mg)	
ACR 20	Proportion of pts achieving ACR 20 at week 24 (%)	0–100	11.7 (v = 11.7)	58.2 (v = 58.2)	59.6 (v = 59.6)	0.074
ACR 50	Proportion of pts achieving ACR 50 at week 24 (%)	0–100	5.8 (v = 5.8)	34.8 (v = 34.8)	36.6 (v = 36.6)	0.129
ACR 70	Proportion of pts achieving ACR 70 at week 24 (%)	0–100	2.4 (v = 2.4)	18.8 (v = 18.8)	16.1 (v = 16.1)	0.185
mTSS	Mean amount of progression of joint damage in hands and feet at week 52 (mean change score ± SD)	0–10	2.8 ± 7.8 (v = 72.0)	0.4 ± 5.7 (v = 96.0)	0.0 ± 4.8 (v = 100.0)	0.185
Infections	Incidence rate of pts experiencing infections and infestations (*n* per 100 subject-years)	70–80	72.13 (v = 78.7)	79.88 (v = 1.2)	76.62 (v = 33.8)	0.037
SAEs	Incidence rate of pts experiencing musculoskeletal and connective tissue disorders (*n* per 100 subject-years)	25–60	57.05 (v = 7.4)	28.39 (v = 79.0)	25.88 (v = 85.3)	0.056
Deaths	Proportion of deaths (%)	0–3	0.15 (v = 95.0)	0.42 (v = 86.0)	0.97 (v = 67.7)	0.185
Tuberculosis	Pts with tuberculosis (*n*)	0–30	0 (v = 100.0)	5 (v = 83.3)	28 (v = 0.067)	0.056
Malignancies	Proportion of pts developing at least one malignancy (%)	0–2	0.9 (v = 23.0)	1.9 (v = 0.2)	1.4 (v = 0.6)	0.093

Reproduced and adapted from the benefit–risk methodology project work package 3 report: field tests [10]

ACR American College of Rheumatology, *MCDA* multi-criteria decision analysis, *mTSS* modified total Sharp score, *pt(s)* patient(s), *SAE* serious adverse event, *SD* standard deviation

0–100 %, whilst the range of 'possible' outcomes for the proportion of deaths in the study was selected taking into account the observed data, in other words from 0 % (best possible outcome) to 3 % (worst possible outcome). The choice of 3 % as the worst outcome (death) avoided eliciting preferences based on worst-to-best swings that are difficult to imagine, such as 100 % of subjects dying during the study.

At this stage, it is useful to structure the criteria hierarchically to create an accurate and analytic understanding of the decision problem. Figure 1 shows a tree first broken down into two fundamental nodes, benefits and risks, which are further divided into sub-nodes. At the end of the tree, we find the model criteria.

2.2 Single Criterion Partial Value Function

Partial value functions $v_i(\cdot)$ are used to re-scale the consequences of each criterion into an interval scale that represents the preferences over the outcomes of that criterion with respect to a range of outcomes used as reference. The model in Table 1 used a 0–100 scale for the value functions, thus a value of 0 was given to the least preferred and a value of 100 to the most preferred consequence of a given criterion. Other scales are possible, such as the 0–1 interval used in the multi-attribute theory of Keeney and Raïffa [15], on which the weighted-sum MCDA model is based.

The EMA BRMP used a linear value function for favourable effects, such as the proportion of subjects with an ACR 20 response, in which a percentage of 11.7 % was simply translated into a value of 11.7. A negative linear function was used for criteria representing unfavourable effects, thus meaning that 0 % of deaths was represented by a value of 100 for the death criterion and 3 % of deaths (worst possible outcome defined for this criterion) would be equivalent to a value of 0.

2.3 Weighting of Criteria

The partial value functions contain no information concerning how important responses in one criterion are with respect to responses in another criterion. As such, weights w_i are used to reflect the relative importance between criteria and are basically scaling factors that relate the importance of a worst-to-best swing in one criterion to a swing in another criterion. Table 1 shows the weights used in the MCDA model on rheumatoid arthritis implemented

by the EMA BRMP scaled so they sum 1. For instance, a swing from 0 to 10 in the modified total Sharp score (mTSS) attribute (defined as the change from baseline to week 52 in the mTSS score) was regarded as important as a swing from 0 to 3 % in the proportion of deaths during the study. In this case example, the swing-weighting method was used to fix the weights [16]. This approach comprises fixing an arbitrary weight, for example 100, for the swing from the worst-to-best outcome for criterion A and asking the decision makers to directly assign a weight for the importance of a worst-to-best swing for criterion B. If the value tree is structured hierarchically, not all criteria need to be mutually compared. In a top-down approach, the higher-level nodes are compared first and the sub-nodes compared in later stages [8–11].

2.4 Primary Results and Sensitivity Analysis

The overall preference value for a therapeutic option is a weighted average of the criteria preference values resulting, for the assessment of Table 1, in the active dose of 200 mg having the best benefit–risk profile, with an estimated value of 55, whereas placebo and an active dose of 400 mg had overall values of 44 and 50, respectively. The reporting of results is not usually limited to these figures, but a broader set of displays is produced, for instance, to visualize the breakdown of the overall scores into their individual components. The PROTECT consortium provides a good overview of visual displays that can be used to report the results of an MCDA model applied for BRA [6].

In the final stage, sensitivity analyses play a critical role in assessing the robustness of the results to alternative but plausible scenarios on the preference weights or the drug's performance. A simple deterministic sensitivity analysis consists of checking how much the overall scores change when we vary certain parameters in Table 1. Visual displays such as the tornado diagram are useful to show the results of a deterministic sensitivity analysis [6].

The MCDA model illustrated in Table 1 has been applied over the clinical trial summaries, assuming they are good approximations of the real parameters. However, as sampling error is an intrinsic characteristic of the clinical trial estimates used to feed this model, statistics such as the confidence interval are of interest in order to carry the uncertainty in the input parameters x_{ij} of Eq. (1) over to the results and distinguish real differences from chance findings. While the weighted-sum model has been said to be unable to account for this variability, and alternative MCDA methodologies such as SMAA have been proposed for this purpose [3, 17], other authors have recommended staying within this framework and extending it beyond the initial deterministic approach [9, 10, 12]. A probabilistic sensitivity analysis requires defining the multivariate probability distribution of the source clinical data summaries to propagate their uncertainty into the overall values via Monte Carlo simulations [18].

3 Validity and Reliability of MCDA for BRA

Validity refers to the extent to which a methodology measures what we hope to measure with no or little bias, whereas reliability refers to the ability to yield consistent results, with minimal variability, over repeated implementations.

Table 2 presents an overview of common challenges faced by MCDA in BRA together with solutions to overcome them. Two frequent problems that occur when identifying attributes, namely the inclusion of preference-complementary and overlapping attributes, are revised in Sects. 3.1 and 3.2, respectively. The validity and reliability of this model rely on the assumption that value functions $v_i(\cdot)$ and weights w_i accurately quantify the preferences of decision makers. Section 3.3 focuses on the inclusion of risk behaviours in the value functions. In Sects. 3.4 and 3.5, two common challenges that arise in the weighting phase are discussed. Finally, problems in performing a fully inferential sensitivity analysis are revised in Sect. 3.6.

3.1 Additive Independence

An important assumption of the weighted-sum of Eq. (1) is that criteria are perfect preference substitutes, also known, in terms of multi-attribute utility theory, as additive independence. However, the additive independence assumption may not hold if decision makers argue that if performance of the ACR response is known to be good, then they might assign a lower preference for a high response in mTss. This assumption is weaker than statistical independence; two endpoints can be statistically correlated but preference independent.

Implementations of MCDA for BRA have acknowledged the inclusion of criteria that may not be perfect preference substitutes; however, the size of this violation has not been routinely evaluated or quantified during the preference-elicitation process [10, 11]. The classical approach to check and measure this violation consists of a trade-off of lotteries [15], although for continuous attributes Delquié and Luo [19] proposed an alternative trade-off that avoids gambles.

Keeping preference-complement criteria in the model over-represents their shared value and thus leads to an upward bias of the overall value. Strong preference-complementary criteria should not be maintained in our model, and we should find a way to combine them into one single criterion. If that is not possible, we may need

Table 2 Common challenges and solutions using multi-criteria decision analysis for benefit–risk assessment

Phase	Challenge	Solutions
Identification of criteria	Lack of additive independence	Check and quantify violations to the additive independence assumption during the preference-elicitation process
		Avoid strong preference-complementary criteria in the model. If not possible, add multiplicative terms
		If mildly preference-complementary criteria are kept, add a sensitivity analysis keeping only preference-substitute criteria
	Overlapping classifications	Avoid classifications that markedly overlap
		If overlapping classifications are kept, add a sensitivity analysis maintaining only non-overlapping classifications
Scoring	Value functions on criteria defined as risk of suffering a certain event	Use primarily linear (risk-neutral) value functions
		Optionally, incorporate specific non-risk-neutral behaviours in the model in a sensitivity analysis
Weighting	Framing effects	Investigate and discuss potential framing effects
		If significant effects are foreseen, frame questions in different ways (sensitivity analyses)
	Trade-offs involving uncertainties	Avoid direct trade-offs involving low probabilities:
		Stepwise approach (mild vs. severe, severe vs. death)
		Use longer time frames
		Replace trade-offs vs. risk of death with trade-offs of life durations
		Adjust probabilities using non-expected utility theory
Probabilistic sensitivity analysis	Correlation among criteria	Avoid independent simulation of criteria
		Use the criteria correlation structure from the patient-level database
		Sub-optimally, perform a sensitivity analysis with artificial, but plausible, correlation matrices

to move away from the simple weighted-sum model into a more complex equation adding multiplicative terms [7]. Moderate preference-complementary criteria may be kept in the primary model but would be better dropped in a sensitivity analysis. The use of hierarchies is helpful here, since weights can be assigned first to the higher-level nodes and then multiplied down to obtain weights for the corresponding lower-level sub-nodes. In this way, we can drop one efficacy criterion while maintaining the overall weight of a higher-level node associated with drugs' efficacy.

On another matter, it is erroneous to exclude important preference-substitute criteria from an MCDA model purely because they are statistically correlated to other model criteria [15].

3.2 Overlapping Classifications

In this section, I examine a closely related phenomenon that arises when we use endpoints such as serious adverse events and deaths that carry overlapping information; any death is, by definition, a serious adverse event. If occurrence of death is one of the criteria in our model, a simple way to avoid double counting entails using non-fatal serious adverse events, rather than any serious adverse events, as an additional criterion to ensure that a subject is only classified in one of the two categories.

A challenge arises when the model is implemented after reporting the data and using the available summaries, since clinical trials tend to be analyzed using overlapping classifications. For example, the EMA BRMP working package 3 acknowledged that double counting was evident in the overlapping definitions of the ACR 20, ACR 50 and ACR 70 criteria. The authors recognized that ACR criteria should have better been defined as non-overlapping segments, but eventually accepted the overlapping definitions used in the input report. All efforts should be made to avoid overlapping criteria in the model. If overlapping classifications remain, a constraint sometimes imposed by the input report, a sensitivity analysis keeping only non-overlapping classifications plays an important role in reassuring that the model results are robust to this violation.

3.3 Non-Linear Value Functions to Model Risk Behaviour

A linear function may not always correctly represent the preferences of decision makers. For example, in clinical trials in which the efficacy endpoint is the mean of a biochemistry parameter achieved at the end of the study, there

may be a threshold at which further increases or decreases in that parameter add no extra benefit and a non-linear value is required [20]. This non-linearity is not intended to include risk aversion in our primary decision model but simply to reflect the intrinsic value of this continuous response.

In this section, the focus is on a very different reason for using a non-linear value function. In decision theory, loss aversion refers to people's tendency to prefer avoiding losses over acquiring gains, whereas risk aversion is a closely related concept that refers to the reluctance to accept a benefit that is uncertain.

Figure 2 illustrates the value function used for the percentage of subjects with malignancies in the MCDA model implemented in rheumatoid arthritis, overweighting the smaller proportions. Assuming that the individual event 'suffering a malignancy during the study' has a certain value representing its importance, the non-linear function defined here over changes in the proportion of subjects who suffer this event incorporates risk aversion into our model. From a descriptive perspective, this non-linearity is consistent with empirical evidence showing that most, but not all, individuals tend to overweight both small probabilities and outcomes that are felt as losses with respect to their status quo [21, 22]. Since the target population may contain both risk-averse and risk-seeking subjects, an MCDA model taking part for one behaviour over the other is debatable and requires justification. A prudent solution here may be to regard BRA as a prescriptive rather than a descriptive task and stay with a straight-forward expected-value approach followed, perhaps, by sensitivity analyses using alternative non-risk-neutral behaviours.

3.4 Framing Effects

During the elicitation process, a framing effect arises when alternative ways of designing the surveys or interviews produce different weights, leading to different decisions [23]. The language used in the questionnaire or the nature of the display might lead to changes in the preference weights.

General framing effects are not the focus of this paper since they exist in any survey instrument; nevertheless, some situations deserve mention since they are common in drug-related BRA. Treatment options where effects are presented as positive outcomes, such as survival rates, may be viewed more favourably than the same options when presented as negative outcomes, such as mortality rates [24, 25]. The mathematical form to communicate risks may also be relevant; questions where the risk of an unfavourable event is framed in terms of frequencies with respect to 100 subjects may produce different weights from those obtained using a ratio in terms of events per 1000 subjects, a phenomenon called ratio bias [26]. Choosing an adequate framing may be especially difficult when preferences are directly elicited from the target population of patients since the importance of certain outcomes, a particular rare event for instance, may be difficult to judge for a patient who has not previously experienced that event.

Although framing effects are always difficult to avoid, this topic warrants a transparent exposition during each implementation of MCDA for BRA. On occasions, only sensitivity analyses might fully reassure on the robustness of the model against differences in the framing of the

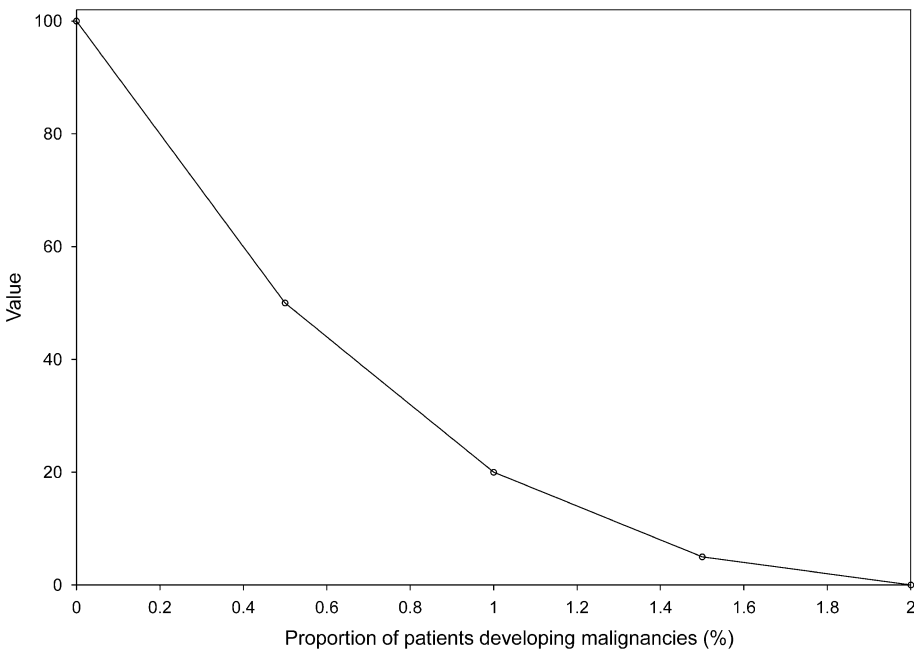

Fig. 2 A non-linear value function over the proportion of malignancies during the study. Reproduced and adapted from the benefit–risk methodology project work package 3 report: field tests [10]

questions used to elicit weights by using questions formulated in different ways, for example.

3.5 Trade-Offs Involving Uncertainties

This section will address a feature specific to BRA not always seen in other applications of MCDA; the frequent need to compare effects with a significant difference in importance. (Generally favourable) mild effects are frequently traded off against the occurrence of (generally unfavourable) severe events that are expected to occur with a low frequency during the study duration [10, 11, 27, 28]. For instance, the MCDA model in rheumatoid arthritis illustrated in Table 1 required the decision makers to compare riskless consequences such as "all subjects will have an ACR 20 response" with other criteria associated with uncertain outcomes such as "3 % of subjects will die during the study." MCDA practitioners frequently use the swing-weighting approach to elicit preferences from decision makers, a method that is primarily meant to compare outcomes defined as certainties [7]. However, an outcome defined as the risk of suffering a particular event during the study, that 3 % of subjects will die during the study for instance, contains (implicitly) a lottery that is typical of the standard gamble (SG) approach, namely the classic method of measuring preferences in economics under uncertainty [29].

Individuals comparing options associated with risk are known to overweight low probabilities and losses over gains [30]. Trade-offs involving uncertainties are particularly insensitive when the interviewee needs to handle low proportions, that is, when there is low probability of a consequence such as death that one is eager to accept for a certain (riskless) improvement in terms of quality of life, for example [21, 22, 31]. Multiple solutions have been proposed to avoid the imprudent use of SG. Bakker et al. [32] recommended avoiding a direct comparison of mildly favourable effects and the risk of events such as death by eliciting these preferences in two steps: an initial comparison of a mild effect with a severe effect, followed by a comparison of the severe effect with death [32]. Johnson et al. [27] suggested using longer timeframes, e.g. 10 years, to increase the magnitude of the expected risks and thus minimize this bias. Bleichrodt et al. [21] proposed the use of descriptive non-expected utility theory to adjust the biased utilities obtained from the SG method into bias-free utilities to be used for prescriptive purposes. However, this method may not remove all the bias and variability caused by elicitations that used probabilities (of death) of less than 20 % [21].

Garcia-Hernandez [33] recently defined a new formulation of the QALYs model that allows transforming utilities obtained using trade-offs of life durations into tolerated excess mortality rates, thereby avoiding direct trade-offs using low probabilities of death during the study.

3.6 Correlation Among Criteria

The rationale for extending the weighted-sum MCDA model into a fully inferential tool through a probabilistic sensitivity analysis was introduced in Sect. 2.4. This section concentrates on a specific feature that makes the task troublesome in this field: clinical trial endpoints are generally positively correlated [34].

Waddingham et al. [35] applied an MCDA model to assess benefits and risks of several drugs in multiple sclerosis and used simulations to capture and propagate the uncertainty of the clinical data. This work initially assumed that criteria were uncorrelated but acknowledged that this approach probably underestimated the variance of the overall benefit–risk values. An additional sensitivity analysis was performed assuming a certain correlation coefficient between all pairs of outcomes. In other research, Wen et al. [36] made use of both the delta method and Monte Carlo simulations to incorporate clinical data uncertainty. The authors used the group-level summaries available in the input report, as well as an arbitrarily taken criteria correlation matrix, and acknowledged that, if the patient-level database were available, resampling methods that obtain the actual correlations among criteria from the clinical database would be more appropriate.

Since this model is essentially a function of a collection of statistics of a very diverse nature, a blocked-by-subject bootstrap method, where one makes the assumption that data are correlated within a subject but independent between subjects, would consist of the following steps:

1. Resample with replacement from the subject-level database (i.e. from a dataset with one record per subject).
2. For each resample, recalculate the summary statistics in the MCDA model (using only the resampled subjects in step 1).
3. For each resample, recalculate the overall benefit–risk score using Eq. (1).
4. Repeat steps 1–3 many times to obtain the bootstrap distribution of the overall benefit–risk scores.

When the patient-level database is unavailable, a sensitivity analysis is still possible using plausible scenarios for the correlations among criteria. Conversely, the independent simulation of criteria ignores their correlation and is, generally, inappropriate.

4 Conclusions

There seems to be a significant gap in the manner in which regulatory authorities and decision makers in other areas such as health technology assessment (HTA) understand

science [37]. The classical reductionist approach to science prevails in the early stages of drug development and regulatory authorities' thinking. Reductionism argues that fundamental principles are difficult to discern within complex systems and warns that more data do not necessarily lead to greater understanding. This approach contrasts with the holistic approach to science that we often encounter in some models used to synthesize evidence for decision making in reimbursement negotiations and health resources allocation. Since many drug development and pharmacovigilance tasks are conducted with the management and close involvement of regulatory authorities, the monitoring and evaluation of benefits and risks have traditionally been conducted under a reductionist paradigm.

Projects such as the EMA-EUnetHTA initiative for a parallel EMA-HTA scientific advice are currently working towards further alignment between regulatory assessments and HTA [37]. Along similar lines of thought, both the EMA BRMP and the PROTECT programme on BRA have explored quantitative methods that could be used to support the continuous benefit–risk monitoring of medicines, eventually selecting the weighted-sum MCDA model for further consideration [6]. Although extensive literature provides case examples, methodological guidance and good practices on the weighted-sum MCDA model in other scientific disciplines, there is currently little experience with this model to appraise benefits and risks of medicinal products [6, 8–11]. This paper has identified potential sources of bias or variability that practitioners may encounter when implementing this model and has also illustrated ways to overcome these challenges.

The weighted-sum MCDA model can be a valid and reliable tool for decision making in BRA provided its implementation adheres to the assumptions of the model with scientific rigour. In order to achieve this objective, a close collaboration between classic stakeholders involved in drug BRA and specialists in the field of decision theory and behavioural economics is essential to ensure that any methodological challenges are routinely identified and adequately addressed.

Compliance with Ethical Standards

Funding No sources of funding were used to assist in the preparation of this study.

Conflict of interest Alberto Garcia-Hernandez has no conflicts of interest that are directly relevant to the content of this study.

References

1. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy—report of the ISPOR Risk-Benefit Management Working Group. Value Health. 2010;13(5):657–66.
2. European Medicines Agency. Benefit-risk methodology project. Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. EMA/549682/2010. 2010.
3. Mt-Isa S, Hallgreen CE, Wang N, Callréus T, et al. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf. 2014;23(7):667–78.
4. European Medicines Agency. Benefit-risk methodology project. Work package 4 report: benefit-risk tools and processes. EMA/297405/2012.
5. Food and Drugs Administration. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making. Draft PDUFA V Implementation Plan—February 2013. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. Accessed 25 Mar 2015.
6. The PROTECT Consortium. Recommendations for the methodology and visualisation techniques to be used in the assessment of benefit and risk of medicines. http://protectbenefitrisk.eu/documents/HughesetalRecommendationsforthemethodologyandvisualisationtechniquestobeusedintheassessmento.pdf. Accessed 25 Mar 2015.
7. Dodgson J, Spackman M, Pearman A, Phillips L. Multi-criteria analysis: a manual. London: Department of the Environment, Transport and the Regions; 2000.
8. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines. Part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepidemiol Drug Saf. 2007;16(Suppl 1):S2–S15.
9. Mussen F, Salek S, Walker S. Benefit-risk appraisal of medicines. A systematic approach to decision making. New York: Wiley; 2009.
10. European Medicines Agency. Benefit-risk methodology project. Work package 3 report: Field tests. EMA/718294/2011. 2011.
11. Nixon R, Dierig C, Mt-Isa S, Stöckert I, Tong T, Kuhls S, Hodgson G, Pears J, Waddingham E, Hockley K, Thomson A. A case study using the PrOACT-URL and BRAT frameworks for structured benefit risk assessment. Biom J. 2015. doi:10.1002/bimj.201300248 **[Epub ahead of print]**.
12. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value Health. 2012;15:1172–81.
13. European Medicines Agency. Assessment report for Cimzia. 2009. Doc.Ref.: EMEA/664021/2009.
14. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, American College of Rheumatology, et al. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum. 1995;38:727–35.
15. Keeney R, Raïffa H. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press; 1993.
16. Von Winterfeldt D, Edwards W. Decision analysis and behavioural research. Cambridge University Press; 1986.
17. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med. 2011;30:1419–28.
18. Broekhuizen H, Groothuis-Oudshoorn CG, van Til J, Hummel JM, IJzerman M. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Parmacoeconomics. 2015. doi:10.1007/s40273-014-0251-x **[ahead of print]**.
19. Delquié P, Luo M. A simple trade-off condition for additive multiattribute utility. J Multi Crit Decis Anal. 1997;6:248–52.
20. Kidney Disease: Improving Global Outcomes (KDIGO) Anemia Work Group. KDIGO clinical practice guideline for anemia in CKD. Kidney Int. 2012;2(Suppl. 2012):279–335.

21. Bleichrodt H, Pinto JL, Wakker P. Using descriptive findings of prospect theory to improve prescriptive applications of expected utility. Manag Sci. 2001;47:1498–514.
22. Van Osch SMC, Wakker P, van den Hout W, Stiggelbout AM. Correcting biases in standard gamble and time trade-off utilities. Med Decis Mak. 2004;24:511–7.
23. Edwards A, Elwyn G. Understanding risk and lessons for clinical risk communication about treatment preferences. Qual Health Care. 2001;10(Suppl I):i9–13.
24. McNeil BJ, Pauker SG, Sox H, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med. 1982;306: 1259–62.
25. Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. Value Health. 2009;12(2):354–63.
26. Pinto-Prades JL, Martinez-Perez JE, Abellán-Perpiñán JM. The influence of the ratio bias phenomenon on the elicitation of health states utilities. Judgem Decis Mak. 2006;1(2):118–33.
27. Johnson FR, Ozdemir S, Mansfield C, Hass S, Miller DW, Siegel CA, Sands BE. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. Gastroenterology. 2007;133(3):769–79.
28. Johnson FR, Van Houtven G, Ozdemir S, Hass S, White J, Francis G, Miller DW, Philips JT. Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. J Neurol. 2009;256:554–62.
29. Gafni A. The standard gamble method: what is being measured and how it is interpreted. Health Serv Res. 1994;28:2.
30. von Neumann J, Morgenstern O. Theory of games and economic behavior. 3rd ed. New York: Wiley; 1953.
31. Van Osch SMC, Stiggelbout AM. Construction of standard gamble utilities. Health Econ. 2008;17:31–40.
32. Bakker C, Rutten M, van Santen-Hoeufft M, Bolwijn P, van Doorslaer E, Bennett K, van der Linden S. Patient utilities in fibromyalgia and the association with other outcome measures. J Rheumatol. 1995;22(8):1536–43.
33. Garcia-Hernandez A. Quality-of-life adjusted hazard of death: a formulation of the quality-adjusted life years model of use in benefit risk assessment. Value Health. 2014;17:275–9.
34. Offen WW, Chuang-Stein C, Dmitrienko A, Littman G, et al. Multiple co-primary endpoints: medical and statistical solutions. A report from the multiple endpoints expert team of the Pharmaceutical Research and Manufacturers of America. Drug Inf J. 2007;41:31–46.
35. Waddingham E, Mt-Isa S, Nixon R, Ashby D. A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment. Biom J. 2015. doi:10.1002/bimj.201300254 **[ahead of print]**.
36. Wen S, Zhang L, Yang B. Two approaches to incorporate clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. Value Health. 2014;17(5):619–28. doi:10.1016/j.jval.2014.04.008.
37. European Medicines Agency Press Office. Early dialogue between regulators and health technology assessment bodies key to medicine development. EMA/738272/2013. 2013.