Highlights

What is already known about the topic?

- MCDA has emerged as a methodology for value-based assessment in health care evaluation and HTA with a number of conceptual, methodological and empirical studies over the recent years.
- Existing evidence from the cognitive psychology literature on whether or not the use of an algorithm is a necessary step for the combination of multiple benefits and their value trade-offs, or whether it can be left to decision-makers' own capabilities, suggests that the human brain has a restricted "integrator" capacity, focusing on a limited number of effects.

What does the paper add to existing knowledge?

- An evidence-based response is provided to a number of claims by the ICER in the US in regards to the use of MCDA methods and stakeholder engagement. These responses relate to aspects of methodological robustness and reliability, complication of use, validation and consensus, and judgement transparency.

What insights does the paper provide for informing health care-related decision making?

- Quantitative decision analysis approaches such as MCDA offer an integration of all relevant benefits and risks for a decision problem, including their value trade-offs, into an overall value function that provides increased scope and depth of the clinically relevant evidence.
- Value preferences can be constructed via decision conferencing, for which hundreds of successful decision analysis applications exist worldwide.

Abstract

The Institute for Clinical and Economic Review (ICER) in the US recently published a 2020 update to its Value Assessment Framework. We are commenting on the method by which the benefits of health interventions are integrated, relating to contextual considerations and other factors relevant to judgments for an intervention's value. We start by discussing the theoretical foundations of decision analysis and its extension to multiple criteria decision analysis (MCDA). Then we provide a detailed, evidence-based response to some of the claims made by ICER in regards to the use of MCDA methods and stakeholder engagement. Finally, we provide a number of recommendations on the use of quantitative decision analysis and decision conferencing that could be of relevance to the ICER methodology before reaching a conclusion. Overall, we agree that some of the proposed changes by ICER are towards the right direction for improving transparency in its value assessment process but that these are probably inadequate. We advocate that more serious attention should be paid on the use of quantitative decision analysis together with decision conferencing for the construction of value preferences via group processes, for the ultimate integration of an intervention's various benefits components.

Introduction

Since 2006, the Institute for Clinical and Economic Review (ICER) in the US has been evaluating the clinical and economic value of innovative health interventions and makes pricing recommendations with the goal of achieving "sustainable access to high value care for all patients"(1). The methodological framework it has developed focuses on the assessment of interventions' long-term value for money and short-term affordability (1). In summary, the ICER value framework adopts a clear conceptual structure as part of which an intervention's long-term value for money is measured by its incremental cost effectiveness, with deliberative adjustments possible for the existence of other benefits and contextual considerations via a majority voting process between members of its appraisal committee. More precisely, in terms of other benefits and contextual considerations, the appraisal committee members are asked to first vote individually on the existence or not of any other benefits or disadvantages and contextual considerations, before being asked to reflect on the voting results as part of the final voting on interventions' long-term value for money. Complementary to the intervention's long-term value for money, its short-term affordability is accounted for via the potential budget impact at health system level, and if in excess of a national threshold (tied to growth of the overall US economy), it acts as a signal for potential access challenges and possibly as a triggering mechanism for policy discussions and negotiation with the manufacturer.

Similar to the public engagement approach adopted by other health care evaluation agencies when seeking to revise their methods, ICER recently invited all interested parties to submit input for the 2020 update of its Value Assessment Framework as part of an open consultation (2). The fourth point of the consultation invitation focused on possible methods by which intervention benefits can be integrated, effectively relating to contextual considerations and other factors relevant when assessing an intervention's value, and possible ways of making these more explicit. Past experience and literature suggest that although similar types of evidence are being assessed during health care evaluation in different countries, the specific evaluation criteria used and the way they are incorporated vary, with their relative importance remaining largely unknown (3). For example, a number of social value elements relating to burden of disease, innovation level, socioeconomic impact and other types of evidence are considered in the majority of cases; however this usually takes place on an ad-hoc basis while lacking transparency regarding the various evidence role and impact, as part of a deliberative decision-making process (3). Such differences in evaluation processes could partly explain the considerable variation observed in drug coverage decisions across countries (4), with evidence suggesting the existence of poor to moderate agreement in HTA recommendations (5). Any significant inter-country

variability in the evaluation processes could have implications on efficiency and fairness that could be argued to risk the reasonableness and credibility of the decisions (6, 7), therefore deserving careful attention. In this paper and in response to the ICER public consultation, we propose an improved way for the incorporation of benefits and other considerations in an effort to inform future updates of its Value Assessment Framework.

A range of different methods exist for the purpose of assessing the value of new health technologies as part of health care evaluation, an interdisciplinary scientific field also known as Health Technology Assessment (HTA). In current HTA practices, a number of value assessment approaches are used across different jurisdictions, which could be broadly divided into (a) comparative clinical benefit assessment, (b) economic evaluation and (c) value-based assessment. We perceive the ICER methodology to pertain to the "value-based assessment" group of approaches having as a mission to serve the needs of decision-makers interested in measuring benefits of interventions that go beyond their clinical value added and cost effectiveness, aiming to capture other aspects of value as part of a more comprehensive approach. This group of approaches is not yet characterized by a single or specific type of methodology, but most approaches have emerged due to the inadequacy of economic evaluation to capture the multi-dimensional nature of new medicines' value in a structured and consistent way. Given that economic evaluation is founded upon the utility-maximising behavior of a single individual, a uniform cost-effectiveness criterion applied at a heterogeneous population level is unlikely to yield Pareto-optimal resource allocations (8). The grounding of cost-effectiveness analysis in von Neumann-Morgenstern's utility theory depends on a set of restrictive utility assumptions including that the Quality Adjusted Life Year (QALY), a measure of health benefit that combines length of life adjusted for quality of life, adequately represents preferences; however this is not always fulfilled (8, 9). This has important implications for the incorporation of utilities at population level, such as equity, fairness and a range of other aspects recently identified by a special ISPOR Task Force Report (10). Examples of such value-based approach efforts include the tiered Incremental Cost Effectiveness Ratio threshold in England based on additional value concerns relating to end-of-life criteria that can effectively increase the valuation of a unit of health outcome for terminal illnesses and also for ultra-rare conditions (11-13), or the fluctuating threshold in Sweden based on disease severity or need (3, 14, 15).

In response to improving consistency in the consideration of additional benefits and the transparency of their influence on drug coverage decisions, multiple criteria decision analysis (MCDA) has emerged as a potential methodology for value-based assessment HTA (16-19). Besides the generation of guidelines for the conduct of MCDA in HTA (20-22), a number of

empirical studies have been conducted in collaboration with decision-makers to test and advance these methods in practice (23-27).

Decision Analysis and MCDA

We believe that the potential usefulness of decision analytic approaches for measuring the overall value of health interventions, including aggregating different benefit components, have not been fully appreciated by ICER. Despite evolving from Expected Utility Theory, these approaches represent an alternative way of thinking to economic evaluation. Perhaps the most relevant theoretical framework relevant to value measurement relates to Decision Theory, with Decision Analysis, its applied discipline, acting as the practical instrument of analysis.

Howard Raiffa, first defined the spirit of Decision Analysis 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) (page 271). Different decision analysis approaches exist which could be broadly divided into qualitative and quantitative types; all approaches contain the definition of Objectives and Criteria, identification of Alternatives and Options, collection of Data and Evidence, and elicitation of Consequences and Preferences. However, quantitative approaches move beyond this to quantify values (or utilities), trade-offs and uncertainty, and aggregate all components together using an algorithm, which can be as simple as a weighted average. Extension of decision analysis applications to include decision problems with multiple objectives led to the foundations of MCDA (29), effectively a fully quantitative decision analysis approach. The development of multi-attribute utility theory from a utility framework is also well illustrated elsewhere (30). The process of MCDA in the context of HTA could be summarized in the phases of problem structuring, model building, model assessment, model appraisal, and development of actions plans (21). In turn, these phases include the steps of selecting and structuring criteria, measuring the performance of alternative options, scoring the options, weighing the criteria, calculating weighted aggregate scores, dealing with uncertainty, and examining the findings (20).

Evidence on whether or not the use of an algorithm is a necessary step for the aggregation of values, trade-offs and uncertainty (e.g. as part of MCDA), or whether it can be left to decision-makers' own capabilities exist from the cognitive psychology literature since the 1950s. In surveying the literature on the construction of complex values, Hastie and Dawes conclude at the end of chapter 11 that "the process of looking first *within* each attribute and then comparing across by some weighting system is superior to that of making global intuitive

judgements *across* attributes" (31). The superiority is due in part to the limitation Miller had noted earlier that a human brain can at one time keep five to nine pieces of information (32). In the behavioral economics literature over the last couple of decades, Thaler and Sunstein raised questions about the rationality of human judgements and decisions due to a number of biases and heuristics(33), followed by Kahneman who more thoroughly acknowledged that the human integrator has restricted capacity (or that the human brain lacks such an "integrator" altogether) (34). More recently, Montibeller and Winterfeldt described the focus on a limited number of effects due to limited mental capacity as "myopic problem representation", explaining in detail 12 cognitive biases and 14 motivational biases that can affect decision makers and experts while making judgements in decision and risk analyses (35). Although the biases focus mainly on modelling uncertainty and values, they also include qualitative considerations associated with problem structuring. For all 26 biases they include guidance about debiasing techniques that can eliminate or reduce them.

Response to the ICER value framework

Although we are not aware of the type of MCDA method(s) considered and tested by ICER together with its independent committees for weighing individual elements, we consider dismissing these methods altogether on the ground that they are not "robust enough to add to reliability of value judgements" (2) (pages 19-20) to be exaggerated. A relevant question to ICER's rejection of MCDA methods would be "not robust compared to what"? As outlined below, there is an increasing accumulation of knowledge that the current process is flawed. At what point in time would the testing of other methods be acceptable?

There have been many successful applications of decision analysis methods across a number of areas, including in drug evaluation (26, 36-41) and other non-health application contexts (42-46). An insightful MCDA application is the SMART Vaccines: Strategic Multi-Attribute Ranking Tool for Vaccines, developed by the Institute of Medicine in collaboration with the National Academy of Engineering (47). From the initial conceptual demonstration of a prioritisation framework to the final empirical application of a software tool for prioritising new vaccines for development, this is a valuable open access resource providing interested users hands-on experience with the concept and use of quantitative decision analysis modelling.

Another relevant MCDA application topic in the context of guiding the development of new health interventions would be along their clinical phases, using information from Target Product Profiles (TPP). A TPP provides a summary format of a new health intervention in development (e.g. drug or vaccine), described in terms of labeling concepts that act as the goal

for the development programme, which is updated dynamically over time (48). For example, the WHO TPPs provide sets of products' attributes reflecting preferred and minimal product characteristics for specific disease areas acting as benchmarks for product development by manufacturers (49). Such information could be used for MCDA model building and assessment, to construct value scales and score the performance of options against the attributes of interest, followed by their weighting. Therefore, in such contexts, well-designed MCDA models could act as mechanisms for guiding the development of future innovations in alignment with what decision-makers value.

This is not to say that the application of MCDA in drug evaluation or HTA comes with no challenges or limitations (50-52). However, it might be more constructive to view any robustness concerns of MCDA studies in alignment with their compliance to good methodological practice (53), something which has been shown to be poor in practice (54), rather than prematurely dismissing this group of methods altogether, leaving no opportunity for their appropriate development and effective application.

Unavoidably, whether these methods are judged to be "too complicated for reliable use" (2) (page 20) or not, will depend on the knowledge, expertise and experience of the people facilitating the overall process, especially during the problem structuring stage the elicitation of value preferences and their trade-offs. In regards to the claimed complexity and time required to build a fully quantitative MCDA model, the Benefit-Risk Project of the European Medicines Agency (EMA) created MCDA models in decision conferences for five new drugs, then under review by the Committee for Human Medicinal Products (CHMP), and easily created each model of the benefit-risk balance within just six hours (55). In terms of their reliability, three separate studies modelling the harms of drugs with different groups of experts produced similar results with high correlations proving a high degree of reliability and accuracy (56). Another early HTA study with two rounds of preference elicitation engaging participants from three different countries resulted in virtually identical results (57). In any case, for some relatively easier decision problems the use of qualitative decision analysis methods (or other methods) might be adequate. Conditions for the choice of quantitative approaches over qualitative ones could be based on the decision's importance and analysis complexity, as for example the severity of the disease indication, the unmet clinical need, the number of outcomes being assessed, the type of trade-offs to be valued and the performance of the treatments.

Similarly, in response to any concerns "that there are no validated or consensus methods to integrate these factors into overall judgements of value", quantitative decision analysis methods are probably the most validated methods known for carrying out such integration of

partial components of value judgements and deriving an overall function of value, as evident through the many theoretical and empirical applications in Keeney's landmark book, Value Focused Thinking (58); in his book, Keeney explains how the overall value of an option is derived based on the extent to which an objective or a number of objectives are judged to add value, and the elicitation of trade-offs that provide a common measure of added value. The implication for any drug evaluation, including HTA, is that although clinical evidence on drug performance for efficacy, safety and quality are based on objective evidence, subjective judgements are always needed for a number of context- and evidence-related concerns, as for example the appropriateness of the data for the intended disease indication, the clinical meaningfulness of the data, or the relative clinical relevance of different benefits and risks. It is in regards to this subjective interpretation of objective data that the use of quantitative decision analysis methods can be of great worth, but also in regards to the overall valuation of evidence spanning both the valuation of performance and the valuation of trade-offs, as they can accommodate these aspects in a structured and transparent way instead of leaving them to become randomly incorporated through ad hoc and vague efforts.

Furthermore, albeit the establishment of cost-effectiveness thresholds, and therefore any "value-based prices", are associated with a number of theoretical and practical challenges if not limitations (59, 60), we are glad to see that one of the ultimate aims of ICER is to "engage all stakeholders in a shared process of learning" in order to "offer a transparent, reliable approach" for integration of benefits (2)(page 23). In this regard, the social psychology literature could be very insightful. In terms of learning processes, it should become clear that preferences do not just "sit in our heads" waiting to be extracted but they need to become constructed in a process of value measurement as part of which "added value" is always a matter of judgement. Construction of preferences can be facilitated through group elicitation processes and it could be argued that "many heads are better than one", as has been illustrated through an experiment on probability distributions obtained from individual versus group-consensus (61). Because of a number of problems relating to interaction processes and cognitive processing, interacting groups (process techniques) might fail to generate judgements as accurate as those of their most capable members, but a combination of group facilitation with judgement analysis and information technology can significantly improve the performance of group's interaction (62). Nevertheless, it should be noted that when making judgements in groups, decision-makers and experts are affected by group-level biases (63). As with the case of individual judgements, group judgements are subject to several biases, with their relative magnitudes depending on factors such as group

size, initial individual judgement, magnitude of bias among individuals, and the group-judgement process adopted (64).

Recommendations on the use of quantitative decision analysis and decision conferencing

The ICER value framework adopts a clear conceptual structure and a well-defined set of benefits, with an incremental cost utility ratio (i.e. incremental cost per QALY gained) acting as the key evaluation metric. Up until recently, the appraisal committee members were following a deliberative voting process for the aggregation of any "other benefits or disadvantages" and "contextual considerations": an initial voting by each member on the existence of any such benefits and considerations, followed by a reflection of the voting results, before a final voting on the interventions' long-term value for money (Figure 1a). The explicit consideration of well-specified "other benefits or disadvantages" (n = 7) and "contextual considerations" (n = 6) is unavoidably a noteworthy feature for improving the transparency of what influences an intervention's long-term value for money.

Following the submission and review of public input to the consultation, ICER published a summary of proposed changes to its 2020 Value Assessment Framework update, rejecting the adoption of any formal multi-criteria decision analytic approach (65). Instead, ICER proposed to retain a "modified approach to integrating other factors into deliberation and decision-making" (65) (page 32), using an expanded set of potential other benefits and contextual considerations for which independent evidence appraisal committee members could vote using a three-level Likert-scale. Following the proposed changes, ICER published the final 2020 updates, as part of which a revised list of other benefits and contextual considerations is considered (n = 9), using the new three-level Likert-scale voting format ranging from Lower Value, to Intermediate Value, to Higher Value (Figure 1b) (66). The goal of this adaptation is to "provide the appraisal committees with a clearer understanding of the end of the spectrum within which they are expected to vote", having also the intention "to produce a more transparent record of how the appraisal committee feels that these considerations should be applied when integrated with the cost-effectiveness results in making decisions about pricing" (66)(page 33). However, it is not fully clear how the various Value levels will inform decision-making; for example, will the Lower Value levels be used to justify no value increment or to justify value decrement, in the form of unchanged or decreased health-benefit (i.e., ex value-based) price benchmarks respectively? Consider the case of the Lower level of "uncertainty in model assumptions" (creating significant risk that the best-case cost-effectiveness estimates are too optimistic), which could be used as a decrement of value. However, the same might not hold true for the Lower Value level of

"mechanism of action" (being similar compared to other active treatments), which could be used to justify a lack of value increment, rather than a decrement.

We agree that the Likert-scale format change is in the right direction for improving the transparency that different factors exert on decision-makers' judgements but it is still inadequate, as transparency requires both access to the underlying performance data and the model structure, in which case it relates both to valuation of performance and valuation of trade-offs. The very last stage of "human integration" required for aggregating together all these additional components into the intervention's core cost-utility ratio in order to derive its' long-term value for money lacks transparency and more importantly is prone to failure and susceptible to bias due to the limited human mental capacity to support such complex tasks, as evident in the behavioral and decision science literature (31, 32, 34, 35, 67). Incorporating additional factors without a clear conceptual framework for the expression of preferences and judgements can degrade the validity of rankings, as has already been demonstrated elsewhere (68).

Among the most important features of quantitative decision analysis approaches and MCDA is the encompassing integration of all relevant benefits for a decision problem, and their value trade-offs, into an overall value function. Recently, the ISPOR Special Task Force on US Value Assessment Frameworks reviewed the use of MCDA for the aggregation of benefits into a single value metric and recommended further research on its use through testing in real-life decision settings as it might provide the best opportunity for improvement (69). In the public policy domain, HM Treasury in the UK recently updated the content of The Green Book regarding guidelines on the appraisal and evaluation of central government projects and programmers to recommend the use of MCDA for estimating the value of social benefits (and costs), including for the consideration of trade-offs (70).

Value preferences could be constructed via decision conferencing, defined as "a gathering of key players who wish to resolve important issues facing their organisation, assisted by an impartial facilitator, using a model of relevant data and judgements created on-the-spot to assist the group in thinking more clearly about the issues" (59) (page 54). Typical stages of decision conference workshops include exploring the issues, structuring and building the model, exploring the model and agreeing on the way forward, all of which can be in alignment with requisite modelling: a decision model whose form and content are sufficient to solve a particular problem (71). Among the requirements for constructive decision conferencing processes are ensuring that a diversity of perspectives is represented and a feeling of "cohesiveness" exists between participants as part of which different opinions are being heard in a trusted manner; group numbers of between 7 and 15 participants have shown to be ideal as they are small enough to

allow participants to reach an agreement but sufficiently large to represent all perspectives and interests (72), which together with appropriate facilitation from a decision analysis expert can engage participants in constructive discourse and peer review which enables the group outputs to be better than even those of the best individual in the group (62). Hundreds of successful decision analysis applications exist worldwide for which preferences have been constructed using decision conferencing (73, 74).

A recent MCDA pilot study in collaboration with European HTA bodies revealed how the application of quantitative decision analysis in combination with decision conferencing could take place, by adopting a clean slate approach without the use of economic evaluation (75). A number of evaluation criteria were incorporated in each jurisdiction's value model and decision-maker value preferences were elicited during decision conferences, including performance scores and criteria weights. Following the estimation of drugs' overall benefits, subsequent consideration of drug costs enabled the demonstration of value for money in the form of estimated "cost per unit of value" ratios. Although findings revealed some differences in value preferences across countries, drug rankings remained consistent. Importantly, the study demonstrated how MCDA could act as a decision support tool for HTA, due to the transparency in the construction of value preferences in a collaborative manner.

In the context of HTA decision-making, assuming that agreement has been reached regarding the level of cost-effectiveness threshold(s), adjusting the threshold or the interventions' incremental cost utility ratio in order to accommodate for other benefits and contextual considerations should be a possible task. For example, as part of an incremental MCDA approach to economic evaluation, a "baseline" threshold could be expanded proportionally with any additional value not captured by the QALY component (76). Alternatively, as part of a clean slate approach independent from economic evaluation, assuming the existence of a well-defined budget for allocation of resources within a particular indication or therapeutic area, a value function could be used together with the purchasing costs of the interventions to calculate multi-dimensional value for money ratios. This would point towards the use of multi-criteria portfolio decision analysis, aiming to maximize benefits given a budget constraint, while allowing for opportunity costs to be naturally incorporated (77). In any case, further research would be required to develop and test such new methodological applications.

In a similar fashion, for the context of ICER's assessments, the selection and aggregation of any relevant "other benefits and disadvantages" and "contextual considerations", including the assignment of relative weights of importance, could be implemented at the level of the independent evidence appraisal committees. Given ICER's use of value-based price benchmarks

representing acceptable long-term value for money, a possible way to fuse together the MCDA results with the use of a baseline incremental cost-utility ratio could be by expanding the latter proportionally to reflect how much of the model's total value is accounted by the non-QALY component (76).

<Figure 1a, b, c>

ICER's past assessment for tisagenlecleucel for Pediatric B-cell ALL

As an illustrative example, take the case of tisagenlecleucel's ICER assessment for the treatment of Pediatric B-cell Acute Lymphoblastic Leukemia (78). In regards to "other benefits and disadvantages", the question was first asked: "Does treating patients with tisagenlecleucel offer one or more of the following other benefits?", followed by the specification of six different benefit aspects (see Table 1). The 13-member evidence committee unanimously (13/13) voted that the "intervention offers a novel mechanism of action or approach that will allow successful treatment of many patients who have failed other available treatments". The majority of the evidence committee (9/13) also voted that the "intervention will have a significant impact on improving return to work and/or overall productivity". Finally, only a minority of the committee voted in favour that the "intervention will significantly reduce caregiver or broader family burden" (4/13), and no member of the committee voted in favour for the existence of any other benefits.

In regards to "contextual considerations", the evidence committee was then asked: "Are any of the following contextual considerations important in assessing tisagenlecleucel's long-term value for money?", followed by the specification of six different contextual aspects (see Table 2). The majority of the committee (12/13) voted that the "intervention is intended for the care of individuals with a condition of particularly high severity", and that there is "significant uncertainty about the long-term risk of serious side effects" (9/13) and that there is "significant uncertainty about the magnitude or durability of the long-term benefits" (9/13). A minority of the committee voted in favour that the "intervention is the first to offer any improvement for patients with this condition" (5/13), that "there are additional contextual considerations that should have an important role in judgments of the value of this intervention" (4/13), and that "the intervention is intended for the care of individuals with a condition that represents a particularly high lifetime burden of illness" (1/13).

Following these two voting procedures, the last voting question was asked corresponding to tisagenlecleucel's overall long-term value for money based on the totality of the evidence:

"Given the available evidence on comparative effectiveness and incremental cost-effectiveness, and considering other benefits, disadvantages, and contextual considerations, what is the long-term value for money of treatment with tisagenlecleucel versus treatment with clofarabine (see Table 3)?". As a result, 7 members voted "intermediate", 3 voted "high" and 1 voted "low", with 2 members abstaining from the voting; panel members that voted "intermediate" or "low" noted that "the high degree of uncertainty regarding long-term benefits and harms" led them to vote for a lower category of value than they would have otherwise, with the two abstaining members also disclosing the same uncertainty issues responsible for precluding their ability to assess long-term value for money.

Based on the above, it becomes evident that the overall 3-step voting process relating to other benefits, contextual considerations and overall long-term value for money is associated with various challenges. Initially, the binary voting on the existence of other benefits and contextual considerations in the first two stages excludes performance valuation, towards understanding the magnitudes of the committee members' value preferences. This could be partially addressed through the use of the recently introduced Likert-scale but still, the relationship between product performance (in terms of benefits) or condition characteristics (in terms of contextual considerations) and value preferences would be characterized by limited granularity. Beyond that, trade-offs valuation relating to the relative importance of the various aspects (pertaining to other benefits and contextual considerations against each other, and versus the core cost per QALY metric) would still remain completely unaccountable for. Ultimately, probably the robustness-determining step of the overall process would be the overly complex integration of all the evidence pieces together in the last stage that also remains completely unfacilitated and which is limited by humans' restricted mental capacity to carry out such tasks.

Assuming that a threshold range of $100,000 - $150,000 per QALY reflects the maximum acceptable long-term value for money based on which health-benefit (i.e. ex value-based) price benchmarks are estimated, $100,000 per QALY could hypothetically be chosen as the baseline threshold for interventions whose value is not associated with any other benefits or contextual considerations. This baseline threshold could then be expanded up to $150,000 per QALY if the intervention fully satisfies all the additional benefit and contextual consideration related value aspects, or a certain number of them deemed to be sufficient for reaching the maximum threshold (Figure 1c). For this to take place, the relative importance of each additional value aspect that is relevant for the particular decision context would need to be elicited as part of each technology's evidence appraisal, as for example using swing weighting via decision conferencing.

Ultimately, the decision of whether an intervention's long-term value for money is deemed acceptable and at what price would be fully transparent and reflected both based on the intervention's performance on an explicit set of evaluation criteria but also based on the relevance and relative importance of this performance as judged by the members of the evidence appraisal committee. However, the details of such a socio-technical process would need much attention and should be clear that MCDA as an approach does not provide an 'off-the-shelf' template. Instead, it must be tailor-made for each field of inquiry, as it has been evident from past experience with EMA and EU HTA bodies, both of which took several years to establish MCDA feasibility tests and methodological developments.

<Table 1>
<Table 2>
<Table 3>

Conclusion

Overall, we propose that the combination of quantitative decision analysis together with decision conferencing be considered by ICER for the purpose of integrating together the various benefit components of interventions through the engagement of different stakeholders following group processes for the construction of value preferences. Ideally, any potential dismissal of MCDA methods by decision-makers and HTA institutions, including health care evaluation agencies like ICER, should be preceded by adequate research in their development and application; for example, this could take the form of case study work involving specific MCDA techniques and the conduct of participatory processes to arrive at the value of interventions for different contexts. We believe that through additional work and improvement by MCDA practitioners and HTA researchers, quantitative decision analysis methods can act as valuable decision-making tools for ICER and other organisations keen in using value-based assessment approaches for HTA. Furthermore, given their explicit and transparent nature, such methods could serve as a template for future innovation, by guiding the development of new health interventions to what the relevant stakeholders, decision-makers and society value mostly. We would therefore recommend for their meticulous hands-on testing via continuous practical applications, in collaboration with the relevant HTA decision-makers, institutions and stakeholders. ICER's mantra is known to be "Fair Pricing, Fair Access, Future Innovation", however fair pricing relies on accurate, consistent and transparent assessment, for which the appropriate construction of value preferences and aggregation of benefits is crucial.

Acknowledgements

The authors are grateful to Dan Ollendorf and four anonymous reviewers for their valuable comments that helped to improve the paper.

References

1. ICER. A Guide to ICER's Methods for Health Technology Assessment. Institute for Clinical and Economic Review, 2018.
2. Institute for Clinical and Economic Review I. Overview of the ICER value assessment framework and update for 2017-2019. https://icer-review.org/wp-content/uploads/2018/03/ICER-value-assessment-framework-update-FINAL-062217.pdf, 2019.
3. Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. European Journal of Health Economics. 2018; 19: 123-52.
4. Clement FM, Harris A, Li JJ, et al. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. JAMA. 2009; 302: 1437-43.
5. Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: A comparative analysis of five countries and implications for coverage decisions. Health Policy. 2012; 108: 167-77.
6. Daniels N, Sabin J. The ethics of accountability in managed care reform. Health Affairs (Project Hope). 1998; 17: 50-64.
7. Daniels N. Decisions about access to health care and accountability for reasonableness. Journal Of Urban Health: Bulletin Of The New York Academy Of Medicine. 1999; 76: 176-91.
8. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. Journal of Health Economics. 1997; 16: 1-31.
9. Pliskin J, Separd D, Weinstein M. Utility Functions for Life Years and Health Status. Operations Research. 1980; 28: 206-24.
10. Lakdawalla DN, Doshi JA, Garrison LP, et al. Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. Value in Health. 2018; 21: 131-39.
11. NICE. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence, 2013.
12. NICE. Developing NICE guidelines: the manual. The National Institute for Health and Care Excellence, 2014.
13. NICE. Interim Process and Methods of the Highly Specialised Technologies Programme Updated to reflect 2017 changes. National Institute for Health and Care Excellence, 2017.
14. Persson U, Willis M, Ödegaard K. A case study of ex ante , value- based price and reimbursement decision- making: TLV and rimonabant in Sweden. Eur J Health Econ. 2010; 11: 195-203.
15. Golan O, Hansen P, Kaplan G, et al. Health technology prioritization: Which criteria for prioritizing new technologies and what are their relative weights? Health policy. 2011; 102: 126-35.
16. Devlin N, Sussex J. Incorporating multiple criteria in HTA: methods and processes. London: Office of Health Economics, 2011.
17. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research. 2012; 15: 1172-81.
18. Kanavos P, Angelis A. Multiple criteria decision analysis for value based assessment of new medical technologies: a conceptual framework. The LSE Health

Working Paper Series in Health Policy and Economics: London School of Economics and Political Science, 2013.
19. Marsh K, Lanitis T, Neasham D, et al. Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. PharmacoEconomics. 2014; 32: 345-65.
20. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research. 2016; 19: 1-13.
21. 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: 435-46.
22. 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 Practices Task Force. Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research. 2016; 19: 125-37.
23. Radaelli G, Lettieri E, Masella C, et al. Implementation of EUnetHTA core Model® in Lombardia: the VTS framework. International Journal Of Technology Assessment In Health Care. 2014; 30: 105-12.
24. Tony M, Wagner M, Khoury H, et al. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. BMC Health Services Research. 2011; 11: 329.
25. Jaramillo HEC, Goetghebeur M, Moreno-Mattar O. TESTING MULTI-CRITERIA DECISION ANALYSIS FOR MORE TRANSPARENT RESOURCE-ALLOCATION DECISION MAKING IN COLOMBIA. 2016; 32: 307-14.
26. Angelis A. Evaluating the benefits of new drugs in health technology assessment using multiple criteria decision analysis: a case study on metastatic prostate cancer with the dental and pharmaceuticals benefits agency (TLV) in Sweden. MDM Policy & Practice. 2018; 3.
27. Angelis A, Linch M, Montibeller G, et al. Multiple criteria decision analysis for HTA across four EU member states: piloting the Advance Value Framework. Social Science and Medicine. 2020; 246.
28. Raiffa H. Decision analysis: Introductory Lectures on choices under uncertainty. Addison-Wesley, 1968.
29. Keeney R, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. 1993 ed. Cambridge: Cambridge University Press, 1976.
30. Ralph LK. Multiplicative Utility Functions. Operations Research. 1974; 22: 22-34.
31. Dawes RM, Hastie R. Rational choice in an uncertain world. San Diego ; London: San Diego ; London : Harcourt Brace Jovanovich, 1988.
32. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review. 1956; 63: 81-97.
33. Thaler RH. Nudge : improving decisions about health, wealth, and happiness. New Haven: New Haven : Yale University Press, 2008.
34. Kahneman D. Thinking, fast and slow. London : Allen Lane ; Penguin, 2011.
35. Montibeller G, Winterfeldt D. Cognitive and Motivational Biases in Decision and Risk Analysis. Risk Analysis: An International Journal. 2015; 35: 1230-51.

36. Nutt DJ, King LA, Phillips LD. Drug harms in the UK: a multicriteria decision analysis. The Lancet. 2010; 376: 1558-65.
37. Angelis A, Montibeller G, Hochhauser D, et al. Multiple criteria decision analysis in the context of health technology assessment: a simulation exercise on metastatic colorectal cancer with multiple stakeholders in the English setting. BMC Medical Informatics and Decision Making. 2017; 17.
38. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. Medical Decision Making: An International Journal Of The Society For Medical Decision Making. 2009; 29: 104-15.
39. Marsh K, Zaiser E, Orfanos P, et al. Evaluation of COPD Treatments: A Multicriteria Decision Analysis of Aclidinium and Tiotropium in the United States. Value in Health. 2017; 20: 132-40.
40. Moore A, Crossley A, Ng B, et al. Use of multicriteria decision analysis for assessing the benefit and risk of over-the-counter analgesics. The Journal of Pharmacy and Pharmacology. 2017; 69.
41. Nixon R. A case study using the PrOACT-URL and BRAT frameworks for structured benefit risk assessment. Biometrical Journal Biometrische Zeitschrift. 2016; 58: 8.
42. Keeney RL, von Winterfeldt D. Managing Nuclear Waste from Power Plants. Risk Analysis. 1994; 14: 107-30.
43. Phillips LD. The Royal Navy's Type 45 story: a case study. In: Salo A, Keisler J, Morton A, eds., *Portfolio*
Decision Analysis: Improved Methods for Resource Allocation. New York: Springer-Verlag, 2011.
44. Gregory R. Using Stakeholder Values to Make Smarter Environmental Decisions. Environment: Science and Policy for Sustainable Development. 2000; 42: 34-44.
45. Bana e Costa CA, Antão da Silva P, Nunes Correia F. Multicriteria evaluation of flood control measures: the case of Ribeira do Livramento. *Water*
Resources Management. 2003; 18: 263-83.
46. French S, Kelly N, Morrey M. Decision conferencing and the international chernobyl project. Journal of Radiological Protection. 1992; 12: 17-28.
47. National Academy of Sciences. SMART Vaccines: Strategic Multi-Attribute Ranking Tool for Vaccines. The National Academies Press, 2015.
48. FDA. Guidance for Industry and Review Staff Target Product Profile — A Strategic Development Process Tool. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/target-product-profile-strategic-development-process-tool, 2007.
49. World Health Organisation. WHO Target Product Profiles (TPPs). WHO, 2020.
50. Morton A. Treacle and Smallpox: Two Tests for Multicriteria Decision Analysis Models in Health Technology Assessment. Value in Health. 2017; 20: 512-15.
51. Marsh K, Sculpher M, Caro JJ, et al. The Use of MCDA in HTA: Great Potential, but More Effort Needed. Value in Health. 2018.
52. Marsh K, Thokala P, Youngkong S, et al. Incorporating MCDA into HTA: Challenges and potential solutions, with a focus on lower income settings. Cost Effectiveness and Resource Allocation. 2018; 16: 43.
53. Phillips LD. Best Practice for MCDA in Healthcare. In: Marsh K, Goetghebeur M, Thokala P, et al., eds., Multi-Criteria Decision Analysis to Support Healthcare Decisions: Springer, 2017.

54. Oliveira M, Mataloto I, Kanavos P. Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art. The European Journal of Health Economics : HEPAC. 2019: 1-28.
55. EMA. Benefit-risk methodology project. Work package 3 report: Field tests. European Medicines Agency, 2011.
56. van Amsterdam J, Nutt D, Phillips L, et al. European rating of drug harms. Journal of Psychopharmacology. 2015; 29: 655-60.
57. Angelis A, Thursz M, Ratziu V, et al. Early Health Technology Assessment during non-alcoholic steatohepatitis (NASH) drug development: a two-round, cross-country, multi-criteria decision analysis. Under peer review. 2020.
58. Keeney R. Value focused thinking: a path to creative decision making. Cambridge: Harvard University Press, 1992.
59. Phillips LD, Bana e Costa CA. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. Annals of Operations Research. 2007; 154: 51-68.
60. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and care excellence cost- effectiveness threshold. Health Technology Assessment. 2015; 19: 1-503.
61. Phillips LD. Group elicitation of probability distributions: are many heads better than one? : Kluwer Academic Publishers, 1999.
62. Reagan-Cirincione P. Improving the Accuracy of Group Judgment: A Process Intervention Combining Group Facilitation, Social Judgment Analysis, and Information Technology. Organizational Behavior and Human Decision Processes. 1994; 58: 246-70.
63. Montibeller G, von Winterfeldt D. Individual and Group Biases in Value and Uncertainty Judgments. In: Dias LC, Morton A, Quigley J, eds., Elicitation: The Science and Art of Structuring Judgement: Springer, 2017.
64. Kerr NL, Maccoun RJ, Kramer GP. Bias in Judgment: Comparing Individuals and Groups. Psychological Review. 1996; 103: 687-719.
65. Institute for Clinical and Economic Review I. 2020 Value Assessment Framework: Proposed Changes. https://icer-review.org/material/2020-value-assessment-framework-proposed-changes/, 2019.
66. ICER. 2020 Value Assessment Framework: Final Framework. Institute for Clinical and Economic Review, 2020.
67. Meehl PE. Clinical versus statistical prediction : a theoretical analysis and a review of the evidence. Minneapolis: Minneapolis : University of Minnesota Press, 1954.
68. Montibeller G, Patel P, Del Rio Vilas VJ. A critical analysis of multi-criteria models for the prioritisation of health threats. European Journal of Operational Research. 2019.
69. Phelps CE, Lakdawalla DN, Basu A, et al. Approaches to Aggregation and Decision Making—A Health Economics Approach: An ISPOR Special Task Force Report [5]. Value in Health. 2018; 21: 146-54.
70. HM T. The Green Book: Central Government Guidance on Appraisal and Evaluation. https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent: HM Treasury, 2018.
71. Phillips L. A theory of requisite decision models. Acta Psychologica. 1984; 56: 29-48.
72. Phillips L, Phillips M. Faciliated Work Groups: Theory and Practice. The Journal of the Operational Research Society. 1993; 44.
73. Dodgson J, Spackman M, Pearman A, et al. Multi-Criteria Analysis: A Manual. 2009 ed. London: Department for Communities and Local Government, 2000.

74. 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.
75. Angelis A, Linch M, Montibeller G, et al. Multiple criteria decision analysis for HTA across four EU member states: piloting the Advance Value Framework. 2020; 246.
76. Phelps C, Madhavan G. Resource allocation in decision support frameworks. Cost Effectiveness and Resource Allocation. 2018; 16.
77. Salo A, Keisler J, Morton A. Portfolio decision analysis : improved methods for resource allocation. New York ; London: New York ; London : Springer, 2011.
78. ICER. Chimeric Antigen Receptor T-Cell Therapy for B-Cell Cancers: Effectiveness and Value, Final Evidence Report. Institute for Clinical and Economic Review, 2018.

Figures and Tables

Figure 1: Previous (a), newly updated (b), and recommended (c) ICER Value Assessment Framework stages focusing on the aggregation of (i) other benefits or disadvantages and (ii) contextual considerations, for (iii) estimating long-term value for money.

1(a)

1(b)

1(c)

Table 1: List of relevant "other benefits or disadvantages" and respective voting results for tisagenlecleucel in the ICER assessment

Table 2: List of relevant "contextual considerations" and respective voting results for tisagenlecleucel in the ICER assessment

Table 3: Voting results for tisagenlecleucel's overall long-term value for money in the ICER assessment