Review

For reprint orders, please contact: reprints@futuremedicine.com

Journal of Comparative Effectiveness Research

Comparison of weighting methods used in multicriteria decision analysis frameworks in healthcare with focus on low- and middle-income countries

Bertalan Németh*,[1], Anett Molnár[1], Sándor Bozóki[2,3], Kalman Wijaya[4], András Inotai[1,5], Jonathan D Campbell[6] & Zoltán Kaló[1,5]

[1]Syreon Research Institute, Budapest, Hungary

[2]Laboratory on Engineering & Management Intelligence, Research Group of Operations Research & Decision Systems, Institute for Computer Science & Control, Hungarian Academy of Sciences (MTA SZTAKI), Budapest, Hungary

[3]Department of Operations Research & Actuarial Sciences, Corvinus University of Budapest, Hungary

[4]Abbott Established Pharmaceutical Division, Basel, Switzerland

[5]Department of Health Policy & Health Economics, Eötvös Loránd University (ELTE), Budapest, Hungary

[6]Department of Clinical Pharmacy, University of Colorado, Aurora, CO, 80045, USA

*Author for correspondence: Tel.: +36 70 466 2508; Fax: +36 1 220 5768; bertalan.nemeth@syreon.eu

Aim: Criteria weighting is a key element of multicriteria decision analysis that is becoming extensively used in healthcare decision-making. In our narrative review we describe the advantages and disadvantages of various weighting methods. **Methods:** An assessment of the eight identified primary criteria weighting methods was compiled on domains including their resource requirements, and potential for bias. **Results:** In general, we found more complex methods to have less potential for bias; however, resource intensity and general participant burden is greater for these methods. **Conclusion:** The selection of the most appropriate method depends on the decision-making context. The simple multiattribute rating technique (SMART) method combined with swing-weighting technique and the analytic hierarchy process methods may be the most feasible approaches for low- and middle-income countries.

First draft submitted: 1 October 2018; Accepted for publication: 14 December 2018; Published online: 15 February 2019

Keywords: criteria weighting methods • low- and middle-income countries • MCDA

Decision-making at micro, meso and macro levels of healthcare is a complex issue, strongly characterized by conflicting aims, that often results in a situation when there is a trade-off between equity and efficiency [1]. Healthcare decisions are therefore ideally not based on a single criterion, but on a set of criteria, such as equity or fairness, efficacy or effectiveness, stakeholder interests and pressures, cost–effectiveness, strength of evidence, and safety [2].

The typical aim of priority setting when considering healthcare interventions is to reach various policy goals; for example, maximizing general population health, reducing health inequalities, supporting disadvantaged or vulnerable groups, while taking into account budgetary constraints. Optimal resource allocation necessitates rational priority setting and a transparent decision-making framework [3], in order to reach health policy objectives. The comparison of different scenarios has to be based on well-defined priorities that reflect on the preferences of the stakeholders. Unfortunately, priority setting when choosing among health interventions is often ad hoc and resources are not used to an optimal extent [3].

Multiple criteria or multicriteria decision analysis (MCDA) – also known as multiple-criteria decision-making (MCDM) – is the collective name of formal approaches that support decision making by taking into account multiple criteria in an explicit and transparent way [4]. Explicitly considering multiple criteria and structuring complex decision problems leads to better understanding of the question, which in turn makes the decision-making process easier (as complex problems are broken down into smaller and more consistent pieces) and more consistent.

Future Medicine

This also implies that the main goals of decision-makers are represented by predefined and clearly stated criteria [5] that optimally result in a transparent decision-making framework [6].

The methodology of MCDA has extensively been used in various disciplines, with healthcare experiencing limited but growing application [7]. The limited use in healthcare so far may be due to several reasons including the complexity of the approach, large data requirements and organizational barriers [6]. It is also important to note that in light of good practice guidelines, most MCDA frameworks have trade-offs between potential methodologic flaws and their ease-of-use [8]. Though a recent evaluation of 41 examples of applying MCDA in healthcare suggested that the favored weighting methods tip toward ease of use over that of potential methodological flaws [9].

Thokala *et al*. listed the following key steps of a value measurement MCDA process in healthcare: defining the decision problem, selecting and structuring criteria, measuring performance, scoring alternatives, weighting criteria, calculating aggregate scores, dealing with uncertainty, and at the end, reporting and examining findings [10]. With the growing need for practical MCDA application, the real-life feasibility of the various methods, which are used to conduct these steps, has great significance. In like manner, the selection of goals, criteria and assessment of the alternatives should be based on a solid methodological background [11].

In our research the focus is on weighting the criteria the step used to quantify the relevance of the selected criteria. As in real-world decision making, various criteria have different levels of importance, which all must be taken into consideration when constructing an MCDA framework. Multiple methods are available to conduct this step, based on different theoretical foundations, and for different levels of required resources [10,12].

Low- and middle-income countries generally have limited resources available for research including a higher discount rate for the future versus present as well as a strictly limited budget [13]. These resource limitations may also have implications for MCDA criteria weight methodology. Prior knowledge of MCDA methods applied to healthcare is often limited in these settings. True for any setting, there is often a requirement from experts and stakeholders to translate theory into practice. The availability of primary data and access to relevant information can also be low. Limited access to software tools, and, in many cases, language barriers, can also potentially discourage researchers from selecting certain MCDA methods.

Though there is high variability in healthcare systems and priority settings across low- and middle-income countries, there is a key area, the field of expired patents, the so-called off-patent pharmaceuticals (OPPs), which are being extensively used in low- and middle-income countries. Low- and middle-income countries require a simple and pragmatic methodology for OPP decision-making that takes more attributes into account than only the price of the products. As a result, building MCDA frameworks to find transparent, consistent and rigorous solutions for decision making for the area of OPPs in these countries has become an emerging practice [14,15].

The objective of this narrative review is to support the construction of MCDA frameworks, by identifying and comparing the most often used criteria weighting methods in order to describe the advantages and disadvantages of the various methods, and to provide guidance for the selection of the most appropriate methodology for a given decision problem in low- and middle-income settings.

Methods

In order to capture the list of the most commonly used criteria weighting methods in our narrative review, we based our selection on the first report of the MCDA Emerging Good Practices Task Force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [10]. The report listed the following approaches: various methods of direct rating, that are generally noncomparative [16], swing weighting [17], the simple multiattribute rating technique (SMART) [18], pairwise comparison methods like the analytic hierarchy process (AHP) [19] or the measuring attractiveness by a categorical based evaluation technique (MACBETH) [20,21], and decomposition methods like discrete choice experiments (DCE) [22,23], the potentially all pairwise rankings of all possible alternatives (PAPRIKA) methodology [24], and conjoint analysis (CA) [25].

Direct rating is possibly the simplest set of methods available for estimating criteria weights, simply performed by asking stakeholders to assign number values to the different criteria, with no trade-offs involved [26].

Setting up criteria levels, based on which the scoring of alternatives is conducted according to the particular criterion, is a type of so-called scoring functions. Swing weighting takes the criteria levels into account when estimating criteria weights. Decision-makers start from a hypothetical worst-case scenario, where all criteria are set to the worst possible levels. Then, they identify the most important criterion by selecting the one that, if improved, would improve the overall situation the most. The most important criterion is assigned 100 points, and then

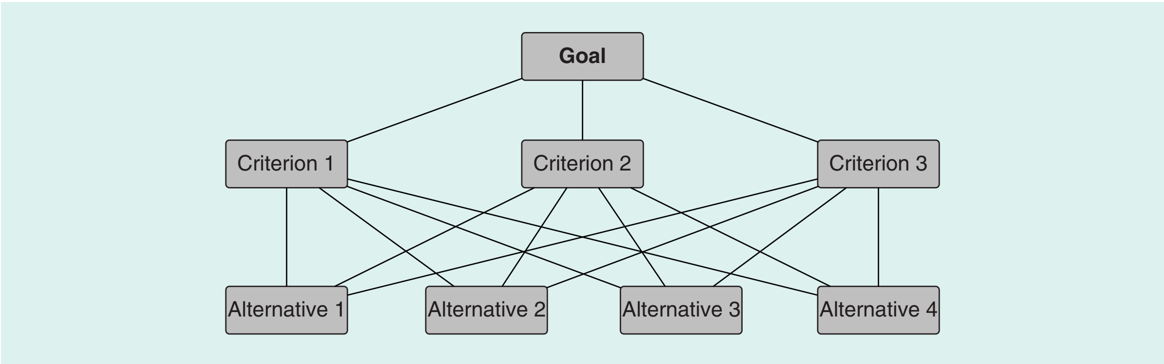

Figure 1. The basic structure of the analytic hierarchy process method (without subcriteria).

decision-makers are asked about the next criterion to be moved from its worst level to its best – that will receive a point value <100 [27].

In case of the SMART method, the weight elicitation is conducted in two phases, during which decision-makers establish an order of importance of the criteria first. Then, starting from the least important criterion, stakeholders assess the relative importance of the next criterion compared with the previously evaluated one by assigning a score to it, higher than the score that was assigned to the previous one. Finally, these weights are normalized to give a sum of 1, or 100% [18]. The SMART and swing methods can be combined as well, using the latter method for setting up the order of importance of the criteria [28].

When conducting an assessment using the AHP, the decision problem is described in a hierarchical, multilevel structure, that is best explained visually in Figure 1, with the goal of making a decision on the highest level. As in the figure, the criteria and subcriteria are placed in the middle, while the alternatives are on the lowest level. This structure attempts to conceptualize the causal relationship concerning how the use of different alternatives could support reaching the main goal. Criterion weights are established based on pairwise comparisons [19]. Pairwise comparisons are used on each level, to provide estimations of the weights for the criteria and for the alternatives as well, resulting in a preference matrix. The criteria weights are determined by the mathematical method of finding the principal eigenvector [29] of this matrix, and then a consistency ratio is defined to assess the consistency of judgements [4]. The eigenvector is the positive vector that changes by only a scalar factor when that linear transformation (expressed as the matrix in question) is applied to it.

With MACBETH, the process is also based on pairwise comparisons across criteria. At the end of evaluation, the M-MACBETH software, that is required to conduct analyses with this method, searches for inconsistency in the resulting matrix of judgments. The software then identifies the minimum number of judgments that should be modified to achieve consistency, and provides suggestions for decision-makers [20].

If a method is based on the DCE principles, the preferences of the decision-maker will only be asked indirectly, using the stated preferences. So-called choice sets are constructed, when decision-makers are faced with two hypothetical scenarios, that have explicitly defined attributes according to various criteria. Generally, decision-makers have to pick the more preferred one from two choice sets, sometimes also with the option to pick 'neither'. Based on their selections, their implicit preferences will be elicited by statistical methods [27]. These methods generally require software support.

The PAPRIKA methodology may be considered the next generation of DCE methodologies. In this case, decision-makers need to choose between two choice sets, as well. During the process, a computer software supports the reduction of the number of questions to answer by assuming transitivity, and excludes pairwise comparisons based on domination. An example of domination is when one alternative has a higher rating on at least one criterion and none lower on any other criterion, compared with the other competing alternative. The PAPRIKA method also provides information on the preferred scoring functions, according to which the alternatives will be assessed by each criterion [24].

It should also be noted, that while some articles used the terms 'CA' and 'discrete choice experiments' interchangeably, the two methods are not the same [30]. DCEs are modeled within a random utility maximization framework.

Table 1. Weighting method trade-off domains.

Method name	Resource requirement	Software requirement	Chance of bias	General complexity
Direct weighting	Low	No	High	Very low
SMART combined with swing weighting	Low	No	Moderate–high	Low
AHP	Moderate	Not necessarily	Moderate	Moderate
MACBETH	Moderate	Yes	Moderate–low	Moderate
DCE	High	Yes	Low	High
PAPRIKA	High	Yes	Low	High
CA	High	Yes	Low	High

AHP: Analytic hierarchy process; CA: Conjoint analysis; DCE: Discrete choice experiments; MACBETH: Measuring attractiveness by a categorical-based evaluation technique; PAPRIKA: Potentially all pairwise rankings of all possible alternatives; SMART: Simple multiattribute rating technique.

While CA evolved from the theory of conjoint measurement, the axioms of which have some relationship to utility theory, and there is also no error theory associated with conjoint measurement [30,31].

A hybrid of adaptive CA is a type of conjoint analyses where decision-makers engage in ranking the criteria, before the pairwise comparison. This method, through software support, can be used to take more attributes and levels of simple scoring functions into account than with other types of CA [27]. With this method, the willingness of decision-makers to trade one attribute for another is taken into account as well as their willingness to pay in the case of the cost criterion.

The methods of weight elicitation listed above were assessed both concerning their theoretical advantages and disadvantages, and also from the point of view of real-life feasibility especially in low- and middle-income countries, focusing on the number and complexity of questions the decision-makers are required to answer. Resource use, such as timescale of developing the MCDA framework, necessary research budget and the number of experts involved, was considered based on the estimation of feasibility of implementation in a setting with limited resources, for example, in low- and middle-income countries. The potential for bias was evaluated based on the scientific literature and the personal experience of authors. The general complexity of the methods was assessed based on the number and complexity of questions, and complexity of calculations.

Results

A brief overview of the findings, as a summary of the authors' judgement, is included in Table 1. The direct weighting method seemingly has a big advantage of being simple, of not needing advanced software and requiring the least amount of resources overall. Although for some decision-makers, conducting pairwise comparisons may be more feasible, the use of this method can also lead to significantly biased results (e.g., participants not expressing their true preferences freely [32]) and some suggested these methods were methodologically incorrect due to their weak scientific basis [33]. Regarding the number of questions, if there are n criteria, using the direct weighting method, decision-makers only have to answer n questions.

Several studies highlighted the main advantages of SMART, which are its simplicity of use, its ease of translating values to a requirement of low resource amounts, and its allowance for both relative and absolute weight assignment techniques [34]. On the other hand, some considered the procedure for determining weight coefficients to be nonconvenient, considering complicated frameworks [35]. Its use is advised when a large amount of information is available and access to decision-makers is easy to obtain [34], and no special software support is available.

Defining criteria weights directly in case of the SMART method combined with swing weighting technique is a simple approach, though it still requires proper explanation. The concern that decision-makers may not be able to quantify the relative importance of their preferences properly is a strong limitation of this approach. Furthermore, as a consequence of the chain structure of ordering the criteria first, there is the possibility that only one incorrectly estimated weight will have a significant impact on the whole structure. Therefore, the potential for bias of this method was classified as moderate–high. The weight of the first criterion may be very small compared with the last question because of the chain effect, especially if there is no cap on the maximum allowed difference between two criteria. With the SMART method combined with swing weighting technique, determining the order of criteria required n-1 questions, and deriving the relative importance by comparison of alternative scenarios required an additional n-1 decisions. This method resulted in a sum of (n-1)*2 questions needed to be answered by decision-makers. Like the previously discussed methods, SMART can also be applied without the use of computers.

The advantage of the AHP method is that it is easy to use, it can adjust in size to accommodate decision-making problems due to its hierarchical structure, and it is straightforward in its approach. AHP's ability to handle larger problems makes it ideal for problems that compare performance among a high number of alternatives [34]. However, several researchers noted that applying this method can potentially lead to inconsistency [36]. In addition, the scales assigning numbers to quantitative judgements of relative importance, which are used during the pairwise comparisons, are arbitrarily set up by the researchers, which can lead to biased results. Among its disadvantages, the interdependence between criteria and inconsistencies was listed (though it can occur between other methods as well). This approach is susceptible to rank reversal, meaning that the addition of alternatives at the end of the process could cause the final rankings to flip or reverse, though resource management problems have a limited number of alternatives to begin with, overall leading to a potential for bias categorized as moderate. With the original AHP approach, seeing that all criteria have to be compared with each other, stakeholders have to make $n*(n-1)/2$ decisions. However, the methodology of incomplete pairwise comparison matrices has been developed [37] and has been applied successfully [38]. This version of the AHP method is able to reduce the number of decisions to be made.

With the AHP method, criteria weights can only be directly compared if decision-makers have used the same factors and/or hierarchies, considering that they can set up different structures individually [39]. In cases where decision-makers have used different hierarchy structures to generate criteria weights, AHP combines their outcomes by taking the geometric mean [40]. Another commonly discussed feature of AHP is that even the decision-makers can be ranked according to their expertise so their individual evaluations can be given more or less importance [41]. The necessity of software support in the case of the AHP method grows with the complexity of the decision problem.

With MACBETH, the requirement of qualitative answers instead of quantitative ones can make using this method easier for several decision-makers, though it can also become a limitation, when discussing criteria that is perceived in a quantitative way in everyday life (e.g., costs). The number of qualitative judgments elicited can range between $n-1$ and $n*(n - 1)/2$, judgements. Though in practice, it is recommended that in addition to these, further judgments should also be elicited with the software [20]. While it requires more resources than the less complex methods, the addition of further judgement elicitation reduces the potential for bias to the moderate–low range.

By using the DCE technique, true preferences can potentially be estimated easier through quantifying the stated preferences of decision-makers. The modeling of competing scenarios provides a more detailed picture of the relative importance of the criteria, which leads to more precise criteria weights. This technique can also be aligned relatively easily with cost-benefit analyses compared with other methods [42]. A report on discrete choice experiments listed potential implausible combinations and interaction effects among the limitations [43]. An additional disadvantage of the method is its potential bias, mainly in cases when potential consequences of bad decisions can be very severe, as some noted when there is an inverse relationship between the magnitude of a decision and the value of applying CA or DCE to it [44]. The main practical issue identified with DCE use was the exponentially increasing number of questions that need to be answered. As the number of combinations of criteria and scoring function levels increase with DCE, so does the number of potential profiles. An increased number of potential profiles can also lead to increasing the number of stakeholder interviews and a potentially large cognitive burden. Furthermore, a method based on DCE principles generally requires software support.

Regarding the number of stakeholders needed for conducting a DCE exercise, it mostly depends on the setting. In practice, the sample sizes of most DCE studies were in the range of hundreds [45]. Most practical designs produce too many questions for a single respondent [43]. A so-called fractional factorial design is often used, where a subset of all possible combinations of attributes are selected, to limit the number of choice sets presented to decision-makers in order to reduce information overload and elicitation burden [46]. The so-called block design can also help to reduce the number of questions [47]. However, care is needed to ensure that the number of choice sets used (combining the various attributes and levels) results in enough data for statistical analysis.

In case of the PAPRIKA methodology, the number of further necessary questions is reduced dynamically as a function of the decisions made previously through eliminating the dominated alternatives. Any number of criteria and/or levels can be included in a survey. However, as the number of criteria and levels increases, the number of potential alternatives (combinations) increases exponentially. The main advantage of the PAPRIKA methodology is that the criteria weights can be constructed without explicitly asking decision-makers about them. Additionally, this technique provides information on the value functions related to the different criteria, which requires serious effort on the part of both the analysts and the decision-makers in the other cases. Naturally, the PAPRIKA method inherits the limitations of the DCE methodology as well. In addition, PAPRIKA also requires the use of particular software.

A study comparing CA with AHP found, that with CA, the scenarios presented to decision-makers were more realistic, while with AHP, participants only compared less concrete attributes, which can lead to less realistic choices, and thus over-, or under-, estimation of the preference weights [48]. A study came to the conclusion that in choice-based conjoint studies around 20 choice tasks can be asked from responders without degradation in data quality [49]. Some argued that CA is generally inconsistent with demand theory in microeconomics and it is unsuitable for use in applied economics due to several logical inconsistencies, while DCEs have a more solid theoretical basis, and are considered to be more general and consistent with demand theory in microeconomics [30]. Just as the case with the DCE and PAPRIKA methods, the CA approach offers a low risk of bias due to its complex approach, but requires more resources and cannot be implemented without software support.

Of the methods assessed, direct weighting, SMART method combined with swing weighting, AHP and PAPRIKA are able to generate individual criterion weights for every decision-maker. With methods such as DCE and CA, estimation procedures such as probit, logit and multinomial logit are used to determine the set of weights for the entire sample [27].

In general, it was also noted that since multiple choice questions are cognitively more challenging, statistically efficient designs may be beyond the reach of respondents with a condition that involves cognitive deficits. The balance between acceptable response efficiency and statistical efficiency may favor simpler designs [43].

Discussion

Several differences were observed across weight elicitation methods, from the methodological point of view, though some methods can be considered minor variants of one another. It is important, however, that in practice their actual usefulness is determined based on procedural and practical considerations [50]. Several recommendations suggested choosing elicitation methods with relatively simple extraction components to be used in applied settings [51]. Some redundancy may even help in getting more robust results, while minimally sufficient sets induce a highly sensitive system of weights.

The required number of questions answered by decision-makers is a considerable limitation for AHP and the general DCE methodologies in the case of more complex decision problems, where the number of criteria is higher. On the other hand, the lower number of required questions for the SMART method combined with swing weighting still carries the risk of inconsistent criteria weights due to the fact that the values of relative importance are explicitly measured. In the study by Mangham *et al.*, up to 18 choice sets were suggested, as 18 is considered a practical limit for how many comparisons can be completed before boredom sets in – though this is likely to vary due to differences in complexity and the characteristics of the target population [52].

A study by van Til *et al.* [53] has reached similar conclusions. In their article, 12 healthy controls and 16 mild cognitively impaired subjects were asked to determine the relative importance of four decision criteria using five different weight elicitation techniques: SMART, SMART combined with swing weighting, Kepner-Tregoe weighting, AHP and CA. In the healthy subjects, no significant preference was found in the feasibility of methods for regular practice, though CA was found to be significantly better regarding general understanding of task, and AHP was significantly the easiest to implement, when considering the difficulty of the task.

In an assessment of weighting methods by Pöyhönen *et al.*, they did not discover essential differences in the results across AHP, the direct weighting method, SMART, swing weighting and tradeoff weighting, though the response scale effects can potentially lead to different weights. The authors in their conclusion suggested that practitioners choose a method following their personal preferences [54].

A study by Mangham *et al.* [52] explicitly discussed the application of the DCE method in a low income country. The authors highlighted the importance of pretesting questionnaires, reducing the number of choice sets reviewed, including pictorial information and verbal descriptions, though some elements of the design can remain complex. The authors also highlighted the importance of communicating the research findings and policy implications in an easily understandable way [52]. Several of these conclusions can be applied to other methodologies as well.

While the different comparisons of weight elicitation methods sometimes showed slightly contradictory conclusions, the most popular methods, which are generally most often used in a setting of countries with limited resources, are the SMART method (possibly in conjunction with swing weighting), AHP and the DCE approach.

The choice between these methods should be based mainly on the availability of resources, including time. This consideration is especially important for low- and middle-income countries that generally have limited resources available for research.

From the methodological point of view, if simplicity is of high importance, or if the number of criteria is relatively high, it is advisable to use the SMART method coupled with swing weighting. In other cases, for example, when different decision-makers need to be weighted based on their relative importance (voting power), then the original AHP method, or incomplete pairwise comparison matrices, can be the most feasible solutions. If flexibility within applied cost–benefit analyses is of key importance, the DCE method can be selected. When adhering to the practice of conducting hundreds of DCE interviews, it can put a strain on researchers, if time and other resources are especially limited. The SMART and AHP methods have been known to deliver results, even after conducting smaller workshops [55].

With the use of these methods, the feasibility of constructing a transparent, consistent and rigorous decision-making framework can be ensured. This idea has been supported by the real-life example of the results of three separate MCDA workshops, which have been conducted to focus on the area of OPPs in low- and middle-income countries, where SMART method combined with swing weighting was used, based on considerations of feasibility [56].

Limitations of our approach include that we based our selection of methods on the weight elicitation methods listed in the first report of the MCDA Emerging Good Practices Task Force of the International Society for Pharmacoeconomics and Outcomes Research [10], and therefore future research could be performed to evaluate more methods or to assess the weight elicitation techniques according to different criteria. We also focused only on the weight elicitation step of an MCDA framework, and did not take into account the other steps, which are also vital parts of particular methods. For example, the technique for order of preference by similarity to ideal solution (TOPSIS) [57] can also be considered, with its unique approach. The main idea is to look for a solution that is as close as possible to the (virtual) positive ideal solution (performing the best with respect to all criteria), and, at the same time, is as far as possible from the (also virtual) negative ideal solution, performing the worst with respect to all criteria. Using one of the two approaches does not necessarily mean that requirements of the other approach are also being fulfilled. MCDA includes both the weighting of the criteria, based on their importance, and applying the scoring functions to the alternatives with respect to each criterion, and the aggregation of the two results in the overall ranking; therefore, separating these can also be considered a limitation of our study.

Conclusion

The results of our narrative review show that there is an inverse relation between the number and complexity of questions to be answered and the complexity of the criteria weighting methodology. Increasing complexity of the criteria weighting methodology is associated with less likelihood for biased findings and higher likelihood of practical challenges. Choosing the most appropriate technique for a certain decision problem has to be done based on the precision tolerance, considering the trade-off between the complexity of questions, the size of the decision-making group and the available resources for constructing the MCDA framework.

Selection between these methods can be based on several specific considerations, like simplicity, the number of criteria or the strength of theoretical basis of the methodologies. In general, pretesting questionnaires, reducing the number of questions that decision-makers need to answer, and communicating the research findings are successful strategies. The SMART method combined with swing weighting technique and the AHP methods can achieve compromises between complexity and the potential for bias, and are likely to be the most feasible approaches for countries with limited resources.

Future perspective

It is very likely that as MCDA is being used in various decision-making environments within the healthcare sector, more information will be available on the real-life usability of the various criteria-weighting methods. This will help in further refining the recommendations on selecting the most appropriate method for a given decision-making context (e.g., low- and middle-income countries with limited resources), taking practical aspects into account.

Executive summary

- Based on a task force report, we identified eight methods for criteria-weighting, which is a key element of multicriteria decision analysis.
- Our selection included general rating methods like the simple multiattribute rating technique (SMART), pairwise comparison methods like the analytic hierarchy process and decompositional methods like discrete choice experiments.
- A qualitative assessment of all reviewed weighting methods was compiled on domains including their resource and software requirements, and their potential for bias.
- In general, we found that methods with higher levels of general complexity may result in less potential for bias; however, resource intensity and general participant burden is greater for these methods.
- The selection of the most appropriate method depends on the real-world context (e.g., availability of resources) in which it is used.
- Low- and middle-income countries generally have limited resources available for research.
- The simple multiattribute rating technique and the analytic hierarchy process methods may be the most feasible approaches for low- and middle-income countries with more limited resources.

Financial & competing interests disclosure

Financial support for this study was provided in part by a contract with Abbott Established Pharmaceutical Division. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing and publishing the report. K Wijaya is employed by the sponsor. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Hsu M, Anen C, Quartz SR. The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* 320(5879), 1092–1095 (2008).
2. Guindo LA, Wagner M, Baltussen R *et al.* From efficacy to equity: literature review of decision criteria for resource allocation and healthcare decision making. *Cost Effectiveness and Resource Allocation* 10(1), 9 (2012).
3. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation* 4(1), 14 (2006).
4. Belton V, Stewart T. *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer Science & Business Media Boston, MA, USA (2002).
5. Broekhuizen H, Groothuis-Oudshoorn CG, van Til JA, Hummel JM, IJzerman MJ. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. *Pharmacoeconomics* 33(5), 445–455 (2015).
6. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health* 15(8), 1172–1181 (2012).

• **Detailed explanation on the use of multicriteria decision analysis in health technology assessment.**

7. Antioch KM, Drummond MF, Niessen LW, Vondeling H. International lessons in new methods for grading and integrating cost–effectiveness evidence into clinical practice guidelines. *Cost Effectiveness and Resource Allocation* 15(1), 1 (2017).
8. Marsh KD, Sculpher M, Caro JJ, Tervonen T. The use of MCDA in HTA: great potential, but more effort needed. *Value Health* 21(4), 394–397 (2018).
9. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. *Pharmacoeconomics* 32(4), 345–365 (2014).
10. 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 Health* 19(1), 1–3 (2016).

• **The first of the two main reports which are used as a starting point of this analysis.**

11. 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* 34(5), 435–446 (2016).

12. 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 Health* 19(2), 125–137 (2016).

- **The second of the two main reports which are used as a starting point of this analysis.**

13. Kaló Z, Bodrogi J, Boncz I *et al.* Capacity building for HTA implementation in middle-income countries: the case of Hungary. *Value Health Regional Issues* 2(2), 264–266 (2013).

14. Brixner D, Maniadakis N, Kaló Z, Hu S, Shen J, Wijaya K. Applying multicriteria decision analysis (MCDA) simple scoring as an evidence-based HTA methodology for evaluating off-patent pharmaceuticals (OPPs) in emerging markets. *Value Health Regional Issues* 13, 1–6 (2017).

15. Brixner D, Kaló Z, Maniadakis N, Kim K, Wijaya K. An evidence framework for off-patent pharmaceutical review (EFOR) for health technology assessment in emerging markets. *Value Health Regional Issues* 16, 9–13 (2018).

16. Bottomley PA, Doyle JR, Green RH. Testing the reliability of weight elicitation methods: direct rating versus point allocation. *J. Marketing Res.* 37(4), 508–513 (2000).

17. von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research* . Cambridge University Press, Cambridge, UK (1993).

18. Edwards W. Use of multiattribute utility measurement for social decision making. In: *Conflicting Objectives in Decisions.*Bell DE, Keeney RL, Raiffa H. (Eds.) John Wiley & Sons, Chichester, UK, 247–276 (1977).

19. Saaty TL. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* 15(3), 234–281 (1977).

20. Bana e Costa CA, Corte J, Vansnick J. On the mathematical foundation of MACBETH. In: *Multiple Criteria Decision Analysis: State of the Art Surveys.* Greco S, Ehrgott M, Figueira JR (Eds). Springer, NY, USA, 421–463 (2005).

21. Ryan M, Gerard K, Amaya-Amaya M (Eds.). *Using Discrete Choice Experiments to Value Health and Health Care.* Springer, Dordrecht, The Netherlands (2007).

22. Kjær T. A review of the discrete choice experiment – with emphasis on its application in health care. Syddansk Universitet, Odense, Denmark (2005).

23. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Economics* 21(2), 145–172 (2012).

24. Hansen P, Ombler F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *J. Multi-Criteria Decision Analysis* 15(3-4), 87–107 (2008).

25. Green PE, Krieger AM, Wind Y. Thirty years of conjoint analysis: reflections and prospects. *Interfaces* 31 (3) S56–73 (2001).

26. Ribeiro F, Ferreira P, Araújo M. Evaluating future scenarios for the power generation sector using a multi-criteria decision analysis (MCDA) tool: the Portuguese case. *Energy* 52, 126–136 (2013).

27. Sullivan T. Using MCDA (multi-criteria decision analysis) to prioritise publicly-funded health care [dissertation]. University of Otago Dunedin, New Zealand (2012). Available at: https://ourarchive.otago.ac.nz/bitstream/handle/10523/2651/SullivanTrudyA2012PhD.pdf?sequence=1&isAllowed=y

28. Edwards W, Barron FH. SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. *Organizational Behav. Human Decision Processes* 60(3), 306–325 (1994).

29. Saaty TL. Decision-making with the AHP: why is the principal eigenvector necessary. *Eur. J. Operational Res.* 145(1), 85–91 (2003).

30. Louviere JJ, Flynn TN, Carson RT. Discrete choice experiments are not conjoint analysis. *J. Choice Modelling* 3(3), 57–72 (2010).

31. McFadden D. Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*. Zarembka P (Ed.) Academic Press, NY, USA (1974).

32. Ishizaka A, Pearman C, Nemery P. AHPSort: an AHP-based method for sorting problems. *Int. J. Prod. Res.* 50(17), 4767–4784 (2012).

33. Bana e Costa CA, Corrêa ÉC, De Corte JM, Vansnick JC. Facilitating bid evaluation in public call for tenders: a socio-technical approach. *Omega* 30(3), 227–242 (2002).

34. Velasquez M, Hester PT. An analysis of multi-criteria decision making methods. *Int. J. Operations Res.* 10(2), 56–66 (2013).

35. Konidari P, Mavrakis D. A multi-criteria evaluation method for climate change mitigation policy instruments. *Energy Policy* 35(12), 6235–6257 (2007).

36. Koczkodaj WW, Szwarc R. On axiomatization of inconsistency indicators for pairwise comparisons. *Fundamenta Informaticae* 132(4), 485–500 (2014).

37. Harker PT. Incomplete pairwise comparisons in the analytic hierarchy process. *Math. Modelling* 9(11), 837–848 (1987).

38. Bozóki S, Csató L, Temesi J. An application of incomplete pairwise comparison matrices for ranking top tennis players. *Eur. J. Operational Res.* 248(1), 211–218 (2016).

39. Bolloju N. Aggregation of analytic hierarchy process models based on similarities in decision-makers' preferences. *Eur. J. Operational Res.* 128(3), 499–508 (2001).

40. Aczél J, Saaty TL. Procedures for synthesizing ratio judgements. *J. Math. Psychol.* 27(1), 93–102 (1983).

41. Saaty TL. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* 1(1), 83–98 (2008).

42. McIntosh E. Using discrete choice experiments within a cost–benefit analysis framework. *Pharmacoeconomics* 24(9), 855–868 (2006).

43. Johnson FR, Lancsar E, Marshall D *et al.* Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health* 16(1), 3–13 (2013).

44. Wainwright DM. More 'con' than 'joint': problems with the application of conjoint analysis to participatory healthcare decision making. *Critical Public Health* 13(4), 373–380 (2003).

45. Salloum RG, Shenkman EA, Louviere JJ, Chambers DA. Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: a systematic review. *Implementation Sci.* 12(1), 140 (2017).

46. Baltussen RM, Ten Asbroek AH, Koolman X, Shrestha N, Bhattarai P, Niessen LW. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health Policy Planning* 22(3), 178–185 (2007).

47. Sandor Z, Wedel M. Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* 38(4), 430–444 (2001).

48. Ijzerman MJ, van Til JA, Bridges JF. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. *Patient* 5(1), 45–56 (2012).

49. Johnson RM, Orme BK. How many questions should you ask in choice-based conjoint studies?. Technical Report, Sawtooth Software 1–23 (1996). Available at: www.sawtoothsoftware.com/technicaldownloads.shtml#howmany

50. Mustajoki J, Hämäläinen RP, Salo A. Decision support by interval SMART/SWING – incorporating imprecision in the SMART and SWING methods. *Decision Sci.* 36(2), 317–339 (2005).

51. Riabacke M, Danielson M, Ekenberg L. State-of-the-art prescriptive criteria weight elicitation. *Adv. Decision Sci.* 2012, Article ID 276584 (2012). www.hindawi.com/journals/ads/2012/276584/

52. Mangham LJ, Hanson K, McPake B. How to do (or not to do)... Designing a discrete choice experiment for application in a low-income country. *Health Policy Planning* 24(2), 151–158 (2009).

53. van Til JA, Dolan JG, Stiggelbout AM, Groothuis KC, IJzerman MJ. The use of multi-criteria decision analysis weight elicitation techniques in patients with mild cognitive impairment. *Patient* 1(2), 127–135 (2008).

•• **A review comparing several criteria weighting methods.**

54. Pöyhönen M, Hämäläinen RP. On the convergence of multiattribute weighting methods. *Eur. J. Operational Res.* 129(3), 569–585 (2001).

•• **A review comparing several criteria weighting methods.**

55. Danner M, Hummel JM, Volz F *et al.* Integrating patients' views into health technology assessment: analytic hierarchy process (AHP) as a method to elicit patient preferences. *Int. J. Technol. Assessment Health Care* 27(4), 369–375 (2011).

56. Inotai A, Nguyen HT, Hidayat B *et al.* Guidance toward the implementation of multicriteria decision analysis framework in developing countries. *Expert Rev. Pharmacoeconomics & Outcomes Res.* 18(6), 585–592 (2018).

•• **Detailed guidance on the selection of methods for low- and middle-income countries.**

57. Hwang CL, Masud AS. *Multiple Objective Decision Making – Methods and Applications: a State-of-the-Art Survey*. Springer Science & Business Media , New York, NY, USA (1979).