Article:
Marsh, Kevin, Sculpher, Mark John, Caro, J. Jaime et al. (2018) The use of MCDA in HTA: great potential, but more effort needed. Value in Health. pp. 394-397. ISSN: 1524-4733

https://doi.org/10.1016/j.jval.2017.10.001

Title: The use of MCDA in HTA: Great potential but more effort is needed

Authors: Kevin Marsh, PhD;[1] Mark Sculpher, PhD;[2] J. Jaime Caro, MDCM, FRCPC, FACP;[3] Tommi Tervonen, PhD[1]

Affiliations:

[1] Evidera, London, UK

[2] Centre for Health Economics, University of York, UK

[3] Evidera, Waltham, MA, US; McGill University, Montreal, PQ, Canada

Corresponding Author: Kevin Marsh, Metro Building, 6th Floor, 1 Butterwick, London, W6 8DL kevin.marsh@evidera.com, Tel: +44 (0)208 576 5025, Fax: +44 (0)208 576 5195

Financial Disclosure: No funding was received to support the writing of this manuscript.

Key words: multi-criteria decision analysis, health technology assessment, reimbursement, good practice

Running title: Critique of MCDA in HTA

Concise summary: Applications of HTA MCDAs often overlook important good practice principles. Awareness of the implications of this practice is required if MCDA is to support HTA.

Acknowledgments: NA

Word count: 2,496

Commented [MK1]: TBD

Number of text pages: 10

Number of tables: 1

Number of figures: 0

Highlights

- *What is already known about the topic?* The potential for MCDA to support HTA has been much discussed, and various HTA agencies are piloting or applying MCDA. Alongside these developments, good practice guidelines for the application of MCDA in healthcare have been developed.
- *What does the paper add to existing knowledge?* An assessment of current applications of MCDA to HTA in light of good practice guidelines reveals that many have methodological flaws that undermine their usefulness.
- *What insights does the paper provide for informing health care-related decision making?* More attention needs to be paid to MCDA good practice by researchers, journal editors and decision makers and further methodological developments are required if MCDA is to achieve its potential to support HTA.

Abstract

The potential for MCDA to support HTA has been much discussed, and various HTA agencies are piloting or applying MCDA. Alongside these developments, good practice guidelines for the application of MCDA in healthcare have been developed. An assessment of current applications of MCDA to HTA in light of good practice guidelines reveals, however, that many have methodological flaws that undermine their usefulness. Three challenges are considered: the use of additive models, a lack of connection between criteria scales and weights, and the use of MCDA in economic evaluation. More attention needs to be paid to MCDA good practice by researchers, journal editors and decision makers and further methodological developments are required if MCDA is to achieve its potential to support HTA.

Introduction

Multi-criteria decision analysis (MCDA) is proposed as part of health technology assessment (HTA) because it offers the means to consider a more comprehensive set of benefits than conventional HTA methods, while still summarising these benefits in a single number. MCDA has been adopted or piloted by various HTA agencies, including in Germany[1], Italy[2], Hungary[3], Colombia[2], and Thailand[4]. While MCDA has the potential to make a contribution to HTA, we are concerned that it is failing to do so because applications are ignoring some key principles. We focus on three challenges: the use of additive models, a lack of connection between criteria scales and weights, and the use of MCDA in economic evaluation. Many of the concerns raised are already recognized in the literature[5,

6]. This commentary highlights and illustrates the implications of these concerns through a discussion of a recent application of a well-known MCDA for HTA[1], the EVIDEM framework[7] (from here on referred to as 'the illustration', see Table 1 for more detail), but the concerns are applicable to many examples of MCDA in HTA.

The Use of Additive Value Models

By far the most prevalent model adopted for HTA MCDAs is the additive one[6]. This involves a simple weighted sum of the criteria scores. An additive model is analytically simple, facilitating implementation and transparency, but it requires that the criteria do not overlap (otherwise there will be inappropriate double counting of value) and are preferentially independent (the weight attached to one criterion should not depend on the performance on other criteria).

Unfortunately, many applications of the additive model in HTA violate these requirements. For instance, in the illustration (Table 1) overlap results from the inclusion of cost-effectiveness alongside other cost and effectiveness criteria. Over 40% of MCDAs designed to support 'coverage or reimbursement' decisions include this overlap[8]. The overlap between these two criteria would be avoided by removing one of them from the MCDA. However, below we advocate removing both of them.

Another source of overlap in the illustration is the inclusion of both 'improvement in patient-reported outcomes (PROs)' and 'improvement in efficacy or effectiveness'. While PROs may provide extra information on outcomes of importance to patients, combining this additively with efficacy or effectiveness requires that the weights attached acknowledge the overlap. The weight attached to 'improvement in PROs' must reflect only the extra value associated with this criterion, over and above that already captured in 'improvement in efficacy or effectiveness'. Eliciting such complex weights would put significant additional cognitive burden on stakeholders. In the presence of such overlap, good practice guidelines recommend re-structuring the criteria set, which can involve re-defining or removing criteria[6]. If the overlap cannot be completely avoided, the MCDA can be run more than

[1] The application was chosen as the most recent publication of the use of EVIDEM available on the EVIDEM website (https://www.evidem.org) at the time of writing the paper. Other applications of the framework may vary in, for instance, the criteria selected and the weighting method adopted.

once, each time just including one of the overlapping criteria, to test whether the different designs generate the same result[6].

Although violation of preferential independence has received less attention in the HTA-MCDA literature, it is perhaps of greater concern, and the broader literature on valuing health provides support for the idea that MCDAs often violate this requirement. For instance, Nord's work on how the value of a particular health gain varies with the underlying disease severity points to a potential preferential dependence of these criteria[9]. However, over 40% of MCDAs designed to support reimbursement or coverage decisions included severity alongside health gain in an additive model[8]. If the value of a health gain varies with disease severity, then its weight will depend on severity level, implying preferential dependence and making an additive model inappropriate.

The illustration shows how preferential dependence undermines the validity of additive models. Consider an intervention, say a placebo that scores zero on 'improvement in efficacy or effectiveness'. Although most would consider such an intervention to have very low value, the additive model can still lead to a high one—a possible overall score up to 0.914 out of 1, depending on performance on other criteria[2]. A placebo with no health gain, but well tolerated and for a severe disease, would be preferred to an intervention with high health gain but minor safety risks for a less severe disease. Concerned to avoid designs that generate such a result, Morton proposes the 'treacle test' for an MCDA—that a completely ineffective treatment should not win[10].

A related source of preferential dependence in the illustration results from the inclusion of both 'comparator intervention limitations' and 'improvement in efficacy'. The value of 'major improvement in efficacy' is dependent on baseline health, which is partly a function of the performance of existing treatment. The illustration puts a similar value on moving from a lower improvement in efficacy to a major improvement in efficacy (0.086) as investing in a treatment whose comparator has major limitations rather than one for which the comparator has no limitations (0.074). However, in combination these criteria produce spurious outcomes because their interaction has not been considered in the design of the MCDA – the model is almost indifferent between investing in a

[2] If an intervention achieved the highest performance on all other criteria except for 'improvement in efficacy or effectiveness', given the weights attached to those criteria, it would in an additive model attain a value of 0.914, even if it receive a score of -3 and thus a partial value of 0 on 'improvement in efficacy or effectiveness'. It is possible that poor performance on 'improved efficacy or effectiveness' might also be reflected in a less than perfect performance on other criteria, such as 'public health interest', 'improvement in PROs' or 'clinical guidelines'. This isn't clear, but the possibility points to sources of overlap in the framework.

treatment that produces little improvement over a comparator that is not effective (0 + 0.074), and investing in a treatment that is a major improvement over a comparator that is already effective (0.086 + 0).

Further work is required to determine what should be done in the presence of preferential dependence. Good practice guidelines would suggest that it should either be removed by restructuring the criteria set and maintaining an additive model, or that the additive model is invalid and a multiplicative or multilinear model should be adopted[6]. These models allow translating zero health gain into an overall zero score, as well as incorporating other interaction effects into the value model. At the heart of the illustration is the idea that issues like this can be dealt with through a deliberative process. We question whether such a process could 'accommodate' a divergence from the analytical assumptions underlying an additive model. Nevertheless, we know of no research on these issues, so further work is required to guide best practice.

No Connection between Criteria Scales and Weights

Weights in additive value models express trade-offs between degrees of performance on the criteria scales. For instance, if a one-year response criterion measured on a scale of 10-90% has a weight of 0.5; while an annual risk of serious adverse events (SAE) measured on a scale of 0-50% has a weight of 0.25, then increasing response from 10% to 90% would be twice as valuable as decreasing annual SAE risk from 50% to 0%. Willingness to trade-off SAEs for clinical benefit will depend on the increase in SAE and the improvement in response being traded. Accordingly, if stakeholders are to provide weights that reflect the trade-offs required by an additive model, they need to do so with scale ranges in mind. Weights should not be elicited independently of the range of consequences[6, 11].

Half of MCDAs where the weighting method was reported adopted weighting methods that ignored the range of consequences of the treatments evaluated[8]. In the illustration, criteria are scored on a pre-defined scale (1-5, with different points on the scale reflecting different levels of importance) and the criterion is defined as, for instance, just 'improvement in efficacy/ effectiveness' between 'lower efficacy/effectiveness' and 'major improvement in efficacy/ effectiveness'. Although use of such a rating scale entails limitations like restricting the relative value of two criteria to a maximum ratio of 5 to 1, here we focus on the challenges posed by not precisely defining the ranges of consequences.

The weight elicitation task adopted in the illustration is the equivalent of posing questions such as: How important is a change from 'lower efficacy' to 'major improvement in efficacy' compared to the change from 'major budget impact' to 'no budget impact'? To answer this, one would need to understand how much health gain is represented by this 'major improvement in efficacy'. In the illustration this is not clear, as the range of consequences is not precisely defined. Nor can it be inferred from the interventions being evaluated, as the weight elicitation is undertaken independently of the details of the interventions, which are revealed to stakeholders only after weights are elicited.

While it is difficult to assess the face validity of weights, which are by their nature subjective, some of the weights elicited in the illustration warrant further scrutiny. For instance, 'type of medical service' received a slightly lower weight (0.067) than 'improvement of safety/tolerability' (0.070). Everything else being equal, this implies that stakeholders would prefer a treatment that offered symptom relief and a major improvement in safety to treatment that was curative but with a worse safety profile. Would stakeholders really not accept some safety risks to cure the disease? That seems unlikely. But, it is not possible to be definitive about this without further information: How severe is the disease being cured? What increased risk of adverse events is to be tolerated? How severe are these adverse events, and for what period are they experienced? Just as it is difficult to face validate the weights without more precision in the definition of criteria, so it is difficult to provide meaningful weights without this extra precision.

It is possible to elicit preferences for precisely defined scale ranges with many different methods. Swing weighting very explicitly ensures that the scales ranges are considered in the elicitation process, as it asks the stakeholders to provide relative importance scores for 'swings' of performance[5]. However, defining precise scales ranges can be challenging when designing a generic MCDA framework to assess technologies that are still in development, as the range of consequences of the treatments to be evaluated is not known at the point at which weights are elicited. Instead it is necessary to either elicit weights across the range of possible performance against each criterion, or to abandon the goal of a generic MCDA framework.

Using MCDA in Economic Evaluation

An important objective of HTA is to estimate the differential benefits and costs of alternative options for particular groups of patients and to present this information in a way that informs funding

decisions. A number of examples of MCDA in the literature have sought to incorporate both the benefits and the costs into a single exercise[8]. For example, EVIDEM's work with 10 medicines in 6 therapeutic areas included both a range of benefit criteria (e.g. improvement of efficacy/effectiveness and disease severity) and economic criteria such as budget impact and cost-effectiveness[12]. This implies that how benefits and costs are ultimately traded-off is determined by the preferences of the stakeholder. There are a two broad concerns raised by this presumption.

First, stakeholders' preferences may not be the appropriate basis for evaluating the efficiency of technologies. Health care systems have agreed funding envelopes, and assuming that there are no obvious examples of unproductive current expenditure, there are always opportunity costs associated with decisions to fund a new and more costly intervention: forgone benefits that could have been generated elsewhere in the health care system. As such, benefit that may be forgone, rather than stakeholders' willingness to pay (WTP), is the most appropriate way of thinking about the incremental cost of the new intervention, and those making funding decisions need to know whether the benefits of the new intervention outweigh the opportunity costs.

Second, even if the goal is to elicit decision makers' WTP, the concerns raised above undermine our confidence in how well this is being done. For example, providing a weight for 'budget impact on the health plan', defined as varying from 'substantial' to 'substantial reduction', seem an almost impossible task. This can be illustrated by interpreting the WTP for health gains implied by the weights in Table 1. Moving from 'substantial budget reduction' to 'substantial budget impact' is given a weight of 0.057. This compares with a weight of 0.086 for 'improved efficacy or effectiveness', moving from 'lower efficacy/effectiveness' to 'major improvement in efficacy / effectiveness". Put another way, this implies that stakeholders are willing to pay 51% more than the difference between a 'substantial budget reduction' and a 'substantial budget impact' to move from 'lower efficacy / effectiveness' to 'major efficacy improvement'. While we can't claim that this is wrong, it is difficult to be confident about, or see how to validate this statement when faced with such imprecise definitions of criteria.

Both these problems are avoided by leaving economic criteria off the list. The potential contribution of MCDA is to reflect the nuances of decision makers' objectives in the form of a more complete benefit measure. The issues dealt with earlier in this paper need to be addressed in developing such a benefit function. Furthermore, any fuller measure of benefit ought to be reflected in terms of an

empirical measure of opportunity costs[13]. Importantly, therefore, MCDA needs to be seen as a means of developing a more complete measure of benefit, hence complementing rather than substituting for efficiency ("cost-effectiveness") analysis.

Conclusion

MCDA has potential for HTA: It can provide a more comprehensive measure of benefit than, for instance, the QALY. Its current application risks undermining this potential, however. We have highlighted three concerns with the approach adopted by many examples of MCDA for HTA: the additive models adopted ignore overlap and preferential dependence between criteria; the simple weighting techniques disregard the ranges of consequence; and the use of MCDA as a form of economic evaluation.

Oftentimes the objectives of 'practicality' or 'proximity' to the thought process of decision makers are raised to justify the approaches employed. It is difficult to see, however, what is practical about adopting an additive model that favours a technology with no health gain. The researcher ought to challenge the decision makers' thinking, helping them understand their objectives and how they relate and interact to generate value, and reflect these in the analytical framework.

The application of MCDA to HTA raises other challenges that are not considered here[14]: Should criteria be fixed or varied between assessments? Which stakeholders should determine criteria and provide weights? How transparent should the criteria and weights be? How should we deal with heterogeneity in weights? It is important that these challenges are not compounded by ignoring established HTA principles and good practice guidelines. We hope this paper increases awareness of these, and what they mean for undertaking MCDA for HTA.

References

1. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods. Available from: https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-2.pdf. [Accessed
2. Castro H, Tringali M, Cleemput I, et al. Advancing MCDA and HTA into Coverage Decision-Making. In: Marsh K, Goetghebeur M, Thokala P, et al., eds., Multi-Criteria Decision Analysis to Support Healthcare Decisions: Springer, Cham, 2017.
3. Endrei D, Molics B, Agoston I. Multicriteria decision analysis in the reimbursement of new medical technologies: real-world experiences from Hungary. Value Health 2014;17(4):487-489.
4. Youngkong S, Baltussen R, Tantivess S, Mohara A, Teerawattananon Y. Multicriteria decision analysis for including health interventions in the universal health coverage benefit package in Thailand. Value Health 2012;15(6):961-970.
5. Marsh K, Thokala P, Muhlbacher A, Lanitis T. Incorporating Preferences and Priorities into MCDA: Selecting an Appropriate Scoring and Weighting Technique. In: Marsh K, Goetghebeur M, Thokala P, et al., eds., Multi-Criteria Decision Analysis to Support Healthcare Decisions: Springer, Cham, 2017.
6. 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 2016;19(2):125-137.
7. Wahlster P, Goetghebeur M, Schaller S, et al. Exploring the perspectives and preferences for HTA across German healthcare stakeholders using a multi-criteria assessment of a pulmonary heart sensor as a case study. Health Res Policy Syst 2015;13:24.
8. 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 2014;32(4):345-365.
9. Nord E, Daniels N, Kamlet M. QALYs: Some Challenges. Value Health 2009;12(S1):S10-S15.
10. Morton A. Treacle and Smallpox: Two Tests for Multicriteria Decision Analysis Models in Health Technology Assessment. Value Health 2017;20(3):512-515.
11. Keeney RL. Common mistakes in making value trade-offs. Operations Reserach 2002;50(6):935-945.

12. Goetghebeur MM, Wagner M, Khoury H, et al. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. Med Decis Making 2012;32(2):376-388.
13. Sculpher M, Claxton K, Pearson SD. Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. Value Health 2017;20(2):234-239.
14. Garau M, Devlin NJ. Using MCDA as a decision aid in Health Technology Appraisal for coverage decisions: opportunities, challenges and unresolved questions. In: Marsh K, Goetghebeur M, Thokala P, et al., eds., Multi-Criteria Decision Analysis to Support Healthcare Decisions: Spinger International Publishing, 2017.

Table 1: Example of the EVIDEM framework (adapted from Wahlster et al. 2015 [7]; https://www.evidem.org/evidem-framework/)

Criterion	Definition	Scale	Norma weight
Relevance and validity of evidence	Extent to which evidence on the proposed intervention is relevant to the decision-making body (in terms of population, disease stage, comparator interventions, outcomes etc.) and valid with respect to scientific standards (i.e., study design etc.) and conclusions (agreement of results between studies).	Low (0) – High (+3) relevance /validity	0.076
Completeness and consistency of reporting	Extent to which reporting of evidence on the proposed intervention is complete (i.e., meeting scientific standards on reporting) and consistent with the sources cited.	Many gaps/inconsistent (0) - Complete and consistent (+3)	0.072
Impact on other spending	Impact of providing coverage for the proposed intervention on other expenditures (excluding intervention cost) such as hospitalization, specialist consultations, adverse events, long-term care, disability costs, lost productivity, caregiver time etc.	Substantial additional other spending (0) - Substantial reduced spending (+3)	0.061
Cost-effectiveness of intervention	Ratio of the incremental cost of the proposed intervention to its incremental benefit compared to alternatives.	Not cost-effective (0) - Highly cost-effective (+3)	0.068

Budget impact on health plan	Net impact of covering the intervention on the budget of the target health plan (excluding other spending). Limited to cost of intervention (e.g. acquisition cost, implementation and maintenance cost).	Substantial budget impact (0) - Substantial budget reduction (+3)	0.057
Type of medical service	Nature of the clinical benefit provided by the proposed intervention at the patient-level (e.g., symptom relief, prolonging life, cure).	Minor service (0) - Major service (+3)	0.067
Public health interest	Risk reduction provided by the proposed intervention at the population-level (e.g., prevention, reduction in disease transmission, reduction in the prevalence of risk factors).	No risk reduction (0) - Major risk reduction (+3)	0.072
Improvement of patient reported outcomes	Capacity of the proposed intervention to produce beneficial changes in patient-reported outcomes (PROs) (e.g., quality of life) compared to alternative interventions.	Worse PRO (-3) - Major improvement (+3)	0.082
Improvement of safety and tolerability	Capacity of the proposed intervention to produce a reduction in intervention-related harmful or undesired health effects compared to alternative interventions.	Lower safety/tolerability than comparators presented (-3) - Major improvement in safety/ tolerability (+3)	0.070

Improvement of efficacy/effectiveness	Capacity of the proposed intervention to produce a desired (beneficial) change in signs, symptoms or course of the targeted condition compared to alternative interventions	Lower efficacy/effectiveness (-3) - Major improvement in efficacy/ effectiveness (+3)	0.086
Comparative interventions limitations	Shortcomings of comparative interventions in their ability to prevent, cure, or ameliorate the condition targeted	No or very minor limitations (0) - Major limitations (+3)	0.074
Clinical guidelines	Concurrence of the proposed intervention with the current consensus of experts on what constitutes state-of-the-art practices	No recommendation (0) - Strong recommendation (+3)	0.067
Size of the population affected by the disease	Number of people affected by the condition (treated or prevented by the proposed intervention) among a specified population at a specified time	Very rare disease (0) - Common disease (+3)	0.063
Disease severity	Severity of the health condition with respect to mortality, disability, impact on quality of life, clinical course (i.e., acuteness, clinical stages).	Not severe (0) - Very severe (+3)	0.080

* A 5-point weight elicitation technique was used (1 = low importance; 5 = high importance). The average of stakeholder weights was used in the MCDA.