ISPOR TASK FORCE REPORT

Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force

Kevin Marsh, PhD[1],[*], Maarten IJzerman, PhD[2], Praveen Thokala, MASc, PhD[3], Rob Baltussen, PhD[4], Meindert Boysen, MSc[5], Zoltán Kaló, MSc, MD, PhD[6],[7], Thomas Lönngren, MSc (Pharm)[8], Filip Mussen, MSc, PhD[9], Stuart Peacock, MSc, DPhil[10],[11], John Watkins, PharmD, MPH, BCPS[12],[13], Nancy Devlin, PhD[14]

[1]Evidera, London, UK; [2]Department of Health Technology & Services Research, University of Twente, Enschede, The Netherlands; [3]University of Sheffield, Sheffield, UK; [4]Radboud Institute for Health Sciences, Nijmegen, The Netherlands; [5]National Institute for Health and Care Excellence, Manchester, UK; [6]Department of Health Policy and Health Economics, Eötvös Loránd University (ELTE), Budapest, Hungary; [7]Syreon Research Institute, Budapest, Hungary; [8]NDA Group AB, UK and Sweden; [9]Janssen Pharmaceutical Companies of Johnson & Johnson, Antwerp, Belgium; [10]Canadian Centre for Applied Research in Cancer Control, British Columbia Cancer Agency, Vancouver, BC, Canada; [11]Leslie Diamond Chair in Cancer Survivorship, Simon Fraser University, Vancouver, Canada; [12]Premera Blue Cross, Bothell, WA, USA; [13]University of Washington, Seattle, WA, USA; [14]Office of Health Economics, London, UK

A B S T R A C T

Health care decisions are complex and involve confronting trade-offs between multiple, often conflicting objectives. Using structured, explicit approaches to decisions involving multiple criteria can improve the quality of decision making. A set of techniques, known under the collective heading, multiple criteria decision analysis (MCDA), are useful for this purpose. In 2014, ISPOR established an Emerging Good Practices Task Force. The task force's first report defined MCDA, provided examples of its use in health care, described the key steps, and provided an overview of the principal methods of MCDA. This second task force report provides emerging good-practice guidance on the implementation of MCDA to support health care decisions. The

report includes: a checklist to support the design, implementation and review of an MCDA; guidance to support the implementation of the checklist; the order in which the steps should be implemented; illustrates how to incorporate budget constraints into an MCDA; provides an overview of the skills and resources, including available software, required to implement MCDA; and future research directions.
Keywords: decision making, healthcare, MCDA, multiple criteria decision analysis.

Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

The task force's first report defined multiple criteria decision analysis (MCDA), provided examples of its use in health care, described key steps, and provided an overview of the principal methods of MCDA [1]. This second task force report provides emerging good-practice guidance on the implementation of MCDA to support health care decisions. Health care analysts who have a basic familiarity with MCDA, but who are not MCDA specialists, are the primary audience for this report. Those concerned about their knowledge of MCDA should consult the first task force report. However, it is hoped that there is good information in the report for MCDA specialists unfamiliar with health care, to support application of their knowledge in a health care context.

Consistent with the first report, this guidance is intended to cover a wide range of decisions, including portfolio optimization, regulatory authorization, health technology assessment (HTA), commissioning decisions/priority setting frameworks, hospital decision making, shared decision making (SDM), prioritizing patients' access to treatment, and disease classification.

An MCDA requires a sociotechnical design, reflecting both the social (who participates, when and how) and technical (which MCDA methods, which software) decisions that need to be made when designing an MCDA [2]. The primary focus of this report is on the technical aspects of the MCDA, though it also provides some guidance on who should be involved in the MCDA and when. The reader is referred to Phillips and Bana e Costa [3] and Bana e Costa et al. [4] for further information on the social

Background to the Task Force

In May 2014, the ISPOR Health Science Policy Council recommended to the ISPOR Board of Directors that an ISPOR Emerging Good Practices Task Force on multiple criteria decision analysis (MCDA) and its use in health care decision making be established. The task force goal was to provide a foundational report on the topic, an MCDA primer, and then focus on initial recommendations on how best to use MCDA methods to support health care decision making.

The task force leadership group is composed of experts in MCDA, health technology assessment, benefit-risk analysis, health care research, pricing, formulary development, epidemiology, and economics. Task force members were selected to represent a diverse range of perspectives. They work in hospital health systems, health technology assessment agencies, research organizations, academia, and the insurance and pharmaceutical industries. The task force had international representation, with members from the United Kingdom, Belgium, Canada, the Netherlands, Sweden, Hungary, and the United States, in addition to reviewers from around the world.

The task force met approximately every 4 weeks by teleconference to develop detailed outlines and discuss issues and revisions. In addition, task force members met in person at ISPOR International Meetings and European Congresses. The four cochairs taught an MCDA course at two of these ISPOR meetings and presented their preliminary findings at workshop and forum presentations multiple times. The final reports were presented at the Third Plenary of the ISPOR 18th European Congress in Milan.

Many comments were received during these presentations. Equally, if not more importantly, both reports were submitted for review twice. Nearly 50 ISPOR members knowledgeable on the topic submitted substantive written comments during these review rounds. All comments were considered. These were discussed by the task force on a series of teleconferences and during a 1.5-day face-to-face consensus meeting. Comments were addressed as appropriate in subsequent versions of the report. We gratefully acknowledge our reviewers for their contribution to the task force consensus development process and to the quality of these ISPOR MCDA task force reports.

All written comments are published at the ISPOR Web site on the task force's Web page: http://www.ispor.org/Multi-Criteria-Decision-Analysis-guideline.asp. The task force report and Web page may also be accessed from the ISPOR homepage (www.ispor.org) via the purple Research Tools menu, ISPOR Good Practices for Outcomes Research, heading: Use of Outcomes Research in Health Care Decisions.

---

elements of implementing the MCDA, such as how to design a workshop to elicit value judgments from participants.

The report's focus is on value measurement approaches because it is our aim to draw emerging good practices from the use of MCDA in health care, and other techniques are rarely used in health care [5]. The first task force report identifies the conditions under which it is appropriate to use value measurement approaches, and alternative techniques that can be adopted when these conditions do not hold.

Value measurement approaches include many techniques (see Table 2). Several of which have been applied in health care [5] and have been the subject of reviews of MCDA methods available to support health care decisions [6–10]. This report draws on this experience. However, given the relative infancy of the application of MCDA in health care, it is also necessary to draw on the broader MCDA literature. (For instance, see Belton and Stewart [11], Guitouni and Martel [12], Velasquez and Hester [13], De Montis et al. [14,15], Getzner et al. [16], Keeney and von Winterfeldt [17], Keeney [18], Dodgson et al. [2], and Olson et al. [19]). The approach adopted in this report is thus to summarize good practice guidance from the nonhealth MCDA literature, interpreting it in light of the characteristics of health care decisions, and referencing health-related examples where they are available.

The terminology and steps adopted in this report are consistent with that in the first task force report. First, although various terms have been used to refer to the value judgments made during an MCDA—for instance, priorities, preferences, importance, values—the reports refer to these judgments as "preferences." Second, the following participants are involved in an MCDA. Decision makers are those who make the choice between alternatives; Stakeholders are the source of scores and weights. The analyst is responsible for the design and implementation of the MCDA. Experts provide advice to other participants on, for instance, the clinical data. Although "stakeholder" is used quite broadly in the health economics literature, within the MCDA literature, the term "stakeholders" is used for those providing the preferences. This terminology is retained in the task force reports. These roles are not mutually exclusive. For instance, depending on the decision problem, the decision maker could also be the stakeholder.

The article is structured as follows. The next section provides the ISPOR MCDA Good Practice Guidelines Checklist to support the design, implementation, and review of the steps involved in an MCDA, and guidance to support the implementation of the checklist. The fifth section (Other Considerations When Designing an MCDA) provides guidance on the order in which these steps should be implemented, and how to incorporate the budget constraints into an MCDA. The sixth section (Resources, Skills, and Software) provides guidance on the skills and resources required to implement MCDA, including the software available to support the implementation of MCDA. The seventh section (Research Directions) summarizes recommendations for the direction of future research.

Good Practice Guidelines

Table 1 presents the ISPOR MCDA Good Practice Guidelines Checklist to support the design, reporting on, and critical assessment of MCDA studies. Given the problem-contingent nature of MCDA methods, the checklist is not intended to be used to prescribe the choice of specific methods. Rather, it provides a list of key considerations when designing and reporting an MCDA. Each step in the checklist includes a recommendation on validation. Following the checklist, general guidance is provided that covers the validation process in each step. Then, detailed guidance is provided on how to implement the other recommendations in the checklist.

Validation

The key role of validation is to confirm that the MCDA design, input, and outputs are plausible and consistent with decision maker objectives and stakeholder preferences. This is especially important given the subjective nature of many of the inputs into an MCDA. Yet, to our knowledge, there is little experience with such validation tasks and we therefore provide only some general recommendations. Validation should be built into each step of the MCDA, and the steps taken to validate the MCDA should be reported. This should include the following:

1. Presentation of the decision problem to decision makers for confirmation.
2. Presentation of the final criteria list and definitions to decision makers, stakeholders, and experts for confirmation. This should consider whether the criteria have the properties

Table 1 – ISPOR MCDA Good Practice Guidelines Checklist.

MCDA step	Recommendation
1. Defining the decision problem	a. Develop a clear description of the decision problem b. Validate and report the decision problem
2. Selecting and structuring criteria	a. Report and justify the methods used to identify criteria b. Report and justify the criteria definitions c. Validate and report the criteria and the value tree
3. Measuring performance	a. Report and justify the sources used to measure performance b. Validate and report the performance matrix
4. Scoring alternatives	a. Report and justify the methods used for scoring b. Validate and report scores
5. Weighting criteria	a. Report and justify the methods used for weighting b. Validate and report weights
6. Calculating aggregate scores	a. Report and justify the aggregation function used b. Validate and report results of the aggregation
7. Dealing with uncertainty	a. Report sources of uncertainty b. Report and justify the uncertainty analysis
8. Reporting and examining of findings	a. Report the MCDA method and findings b. Examine the MCDA findings

MCDA, multiple criteria decision analysis.

required, both as of a set of criteria and as individual criterion (see steps 2a and 2b).
3. Presentation of the performance matrix (see step 3a) to decision makers and experts for confirmation.
4. Testing the consistency of scores and weights through the following:
    a. Eliciting stakeholders' reasons for their preferences. This will allow the analyst to test whether stakeholders' understanding of elicitation tasks is consistent with how their responses will be used.
    b. Consistency checks. The analyst should either report back to stakeholders their interpretation of their preferences for confirmation—for instance, identify changes in criteria that have the same value [2,20]—or elicit preferences multiple times to test the consistency of responses (for instance, Goetghebeur et al. [21]).
5. Presentation of the results of the MCDA to stakeholders for confirmation, drawing attention to the trade-offs that are being made in the MCDA to arrive at these results.

1a. Develop a Clear Description of the Decision Problem
The appropriate MCDA approach will depend on the decision problem. The first step in designing an MCDA should be to develop a clear description of the decision problem, including decision-makers' objectives, including whether the objective is to rank or value alternatives; whether the decision is one-off, or whether a reusable model is required—one that will be used across multiple decisions; alternatives; stakeholders; and decision constraints, such as budgets. Preferences should be provided by the stakeholders whose value judgments are relevant to the decision

problem. The first task force report provides examples of stakeholders who may be relevant for different decisions [1].

When defining the decision problem, the analyst should consult widely with decision makers, experts, and stakeholders and review previous decisions. Tools are available to structure the definition of the decision problem, such as the Criteria, Alternatives, Stakeholders, Uncertainty, and Environment (CAUSE) checklist [11]. Franco and Montibeller [22] review problem structuring tools available to support this task and conclude that although the field of problem structuring methods is well established in management science, more research is required to tailor these tools for use in MCDA. Many health decision makers already acknowledge the importance of this step and invest significant resources in defining the decision problem, such as the scoping process undertaken by reimbursement and regulatory agencies. In these circumstances, the benefit of using tools to support problem structuring may be marginal.

2a. Report and justify the methods used to identify criteria
Decision criteria are the factors that are considered relevant to the decision. Task force report 1 provides examples of the types of criteria relevant to different types of health care decisions. Criteria can be identified in several sources, including documents describing previous decisions; evaluations to support related decisions; studies of stakeholders' priorities; and treatment guidelines. For instance, when undertaking MCDA to support HTA, a wealth of existing material can be drawn on, including reports of previous decisions; decision-making guidance provided by the HTA agency (such as National Institute for Health and Care Excellence's Guide to the Methods of Technology Appraisal [23]); research used to inform these guidance (such as National Institute for Health and Care Excellence's work on social value judgments [24]); reviews of factors considered by HTA agencies (for instance, Youngkong et al. [25]); and surveys of health care decision makers (for instance, Tanios et al. [26]). Consultations with decision makers, stakeholders, and experts will usually form part of criteria identification. The analysts should be aware of the potential biases that may invalidate such consultation, and techniques that can be used to mitigate the impact of these biases (see Montibeller and von Winterfeldt [27]).

The first stage of criteria identification usually results in a long list of potential criteria. This should then be shortened by the analyst in correspondence with the properties required of a set of criteria [2]:

1. Completeness: The criteria should capture all factors relevant to the decision.
2. Nonredundancy: Criteria should be removed if they are unnecessary or judged to be unimportant. For instance, when the objective is to rank alternatives as part of a one-off decision, if alternatives achieve the same level of performance on a criterion, that criterion could be considered redundant. This will avoid stakeholders having to score and weight a criterion that will not have an impact on the results of the MCDA. However, this efficiency gain should be offset against the potential concern of decision makers if a key objective is excluded from the analysis [28]. If this risks undermining the credibility of the analysis, it may be preferable to include the criterion and demonstrate that it does not affect the choice of alternative.
3. Nonoverlap: Criteria should be defined to avoid double counting and thus to avoid giving too much weight to a value dimension. For instance, the assessment of treatments for psoriatic arthritis often use the American College of Rheumatology 20% and 50% improvement criteria scales—the proportion of patients achieving 20% and 50% improvements in seven measures of disease severity. However, including both scales in an MCDA would double count the patients achieving a 20%

improvement in symptoms, who are captured by both measures. In this instance, variation in performance on these two criteria, as well as differences in preferences for the criteria, may mean that the result of the MCDA will vary depending on which criteria is included, and so both criteria may need to be tested in the model separately. Other examples of overlap include discontinuation events and safety events in the same analysis, if discontinuation events may be caused by the safety events; and including cost-effectiveness as a criterion alongside cost and/or effectiveness criteria. It is important that overlapping is not confused with correlation. Criteria can be correlated while still measuring separate objectives.

4. Preference independence: When applying an additive model (see step 6a), how much one cares about the performance on a criterion should not depend on the performance of other criteria. That is, additive models do not allow for the interaction between criteria [29]. Including separate criteria for health gain and severity of disease may violate this requirement because the preference for a gain in health may depend on baseline health. Nord et al. [30] identify the neglect of the interaction between health gain and baseline severity as one of the critiques of how quality-adjusted life-years are used in cost-utility analysis. Using an additive model in the presence of such interactions potentially generates the counterintuitive result of giving a positive value to an alternative that generates no health gain [31].

Two other commonly used criteria that also violate this requirement are frequency and mode of administration. The preference for increased frequency of administration will depend on the mode of administration—adding one pill is not likely to be as bad as adding another injection.

Failure of preference independence can be either realized when the criteria are being formed, or discovered when scoring the alternatives, when stakeholders say they cannot judge their preference for one criterion without knowing scores on the other criterion [2]. In such circumstances, criteria can be redefined to correspond with the requirements of addition models [2,32]. For instance, dependent criteria can be combined into a single criterion—frequency and mode of administration can be combined into a single criterion with levels such as "pill twice a day" and "injection twice a week." Alternatively, multiplicative functions for aggregating criteria can be adopted (see step 6a).

The use of value trees can support the identification of criteria (e.g., see Mt-Isa et al. [33]). A value tree decomposes the objective of an evaluation into subobjectives, organizing them into a hierarchy by clustering them into higher-level and lower-level objectives (see Berkeley and Humphreys [34], Stillwell and von Winterfeldt [35], von Winterfeldt and Fasolo [36], and Hughes et al. [37]). Franco and Montibeller [22] review tools available to generate value trees. These are broadly organized into two types —top-down (using "value-focused thinking" to identify fundamental objectives, and decomposing these into subobjectives, for instance, by asking "how do you achieve that?") and bottom-up (identifying characteristics that distinguish alternatives, which are grouped to form higher-level objectives).

Top-down and bottom-up approaches can generate different results, so the choice of approach is important [38]. Top-down approaches generate sets of criteria that are fairly general, but may be difficult to relate to a particular alternative. Bottom-up approaches produce sets of criteria that are very specifically relevant to the problem at hand. Top-down approaches may, thus, be more appropriate to identify criteria for reusable models, and bottom-up approaches may be more appropriate for one-off decisions.

There is no rule as to how many criteria should be included in an analysis. A recent review of MCDAs in health care found that an average of 8.2 criteria were used to assess interventions, with

the number of criteria ranging from 3 to 19 [5]. It is good practice to have as few criteria as is consistent with making a well-founded decision, though the analyst should consider the trade-off between an increase in validity from a more complete set of criteria and the potential for reducing the validity of scores or weights as a result of the time and cognitive effort associated with more criteria. For instance, inconsistency in patient's responses to pairwise comparison methods have been attributed to fatigue resulting from the length of the questionnaire used [39]. Similarly, large numbers of attributes in a discrete choice experiment (DCE) can be difficult for respondents to process, and most DCEs in health care use four to five attributes [40].

2b. Report and justify the criteria definitions

Once criteria have been identified, they should be defined. Individual criteria should have the following properties: unambiguous (clear relationship between the impact of an alternative and the description of the impact), comprehensive (covering the full range of possible consequences), direct (describe as directly as possible the consequence of implementing an alternative), operational (the information required by the criterion is available and it is possible to make value trade-offs), and understandable (consequences of the criterion are clearly understood by decision makers) [22]; see also Keeney [41].

Direct criteria require that, where possible, proxy outcomes be avoided in favor of "fundamental objectives" [18,22,27]. Fundamental objectives state the reason we are interested in a decision problem (e.g., reducing stroke), whereas a proxy outcome would include intermediate variables (e.g., reducing blood pressure). It is easier to elicit stakeholders' preferences for fundamental objectives. Eliciting preferences for proxy outcomes leaves stakeholders with the challenge of considering how fundamental objectives will be affected by the change in the proxy. Fundamental objectives can be arrived at by repeatedly asking decision makers and stakeholders, "Why do you care about that?"

Criteria are easier to operationalize if they are defined in terms of absolute scales, rather than change estimates such as odds ratios, because preferences for change estimates require knowledge of the baseline value [10]. For instance, preferences for halving the risk of experiencing a serious adverse event will depend on what the risk was beforehand. Operationalizing criteria can also be supported by adopting natural scales over constructed scales [22]. Natural scales are in general use and have a common interpretation. Constructed scales are created specifically for the analysis.

Objective scales are easier to operationalize because they distinguish the "factual" performance measurement from the value judgments involved in scoring and weighting. Most MCDA methods are capable of combining different types of performance measures: quantitative scales, based either on objective (e.g., probability of experiencing an adverse event) or on subjective (e.g., patient-reported outcomes) criteria, alongside qualitative scales (e.g., Tony et al. [42] use qualitative scales to incorporate service capacity and political context into an MCDA designed to support HTA).

The range required for a criterion to be comprehensive will depend on several factors. First, whether criteria will be applied as part of a reusable model. Some decisions require that criteria and preferences be applied consistently across multiple decisions and can benefit from reusable models, including HTA, prioritizing patients, and some commission and prioritization frameworks. Others involve decision-specific models. Regulatory decisions need to be made consistently, but will involve decision-specific criteria. The use of criteria over multiple decisions will require that the range cover the best and worst performance that could realistically occur ("global" or "fixed" scales). Alternatively, where criteria will be applied to a one-off decision, the range can simply

reflect the best and worst performance observed with the alternatives being evaluated ("local" or "relative" scales). Second, how uncertainty is addressed in the MCDA (see step 7a). This may require the range to cover the possible variation in performance, with a rule of thumb being to use a range that includes the 95% confidence intervals of the range of performance of alternatives on the criteria [10]. Third, it is important to avoid a range that exceeds stakeholders' experience, which will raise challenges with eliciting scores and weights [20].

3a. Report and justify the sources used to measure performance

Once the criteria are identified and defined, it is necessary to measure the performance of alternatives—the collection and synthesis of data to assess alternatives on each criterion. The method for measuring performance should conform to the broad principles of evidence-based medicine (see, for instance, Busse et al [43] and [44]) and to local methods guidelines (see, for instance, National Institute for Health and Care Excellence [23]). Often such guidelines will recommend analysis of trial data or network meta-analysis to generate evidence on performance. These often report relative effect estimates, which will need to be translated into absolute values by combining them with reliable estimates of baseline effect (Tervonen et al. [10] demonstrate this for an MCDA designed to evaluate statins). Trial or network meta-analysis data will not always be available to inform performance measurement. This has been identified as a challenge by authors of MCDAs in health care, in particular for criteria such as disease severity, longer-term economic impact, and the feasibility and acceptability of alternatives [5]. In this case, expert opinion should be used to fill the data gap and the impact of the uncertainty in these data should be explored (see step 7a).

The results of the performance evaluation should be displayed in a performance matrix, showing the performance of each alternative against each criterion (see European Medicines Agency [45] for examples of performance matrices). This should include estimates of average performance, variance in this estimate, and the sources of data.

4a. Report and justify the methods used for scoring

The objective of scoring is to capture stakeholders' strength of preferences for changes in the performance *within* a criterion. Scores differ from performance measures in two ways. First, the scores translate performance measures onto a common scale. A 0 to 100 scale is often used to generate scores. This has the advantage of avoiding using decimals, which may be required for shorter scales, and avoids potential confusion with probabilities, which could happen if a 0 to 1 scale is used [20]. Second, scores incorporate strength of preferences for difference in performance.

It is important to clearly explain to stakeholders what performance levels the ends of scoring scales refer to because these reference points will impact the interpretation of scores and weights [20]. For instance, if the "0" on the scoring scale corresponds with zero performance (a ratio scale), a score of 100 should have a value twice that of a score of 50. This property can be used when assessing the consistency of stakeholders' responses (see the "Validation" section). This is not the case when the "0" on the scoring scale does not correspond with zero performance. In this instance, the difference in scores is the basis for consistency checks, using a question such as "Is a change in score of 40-80 really twice as good as a change of 20-40?" To illustrate, if a criterion has a range of 6 (scored 0) to 10 (scored 100), and we assume a linear partial value function, we can say that going from 6 to 8 (0–50) is as good as going from 8 to 10 (50–100), but we cannot say that 10 is twice as good as 8.

Table 2 summarizes a typology of scoring and weighting methods. This covers the stated preference approaches used by most of the MCDAs undertaken in health care. Alternatively, revealed preference approaches could be used to estimate decision makers' preferences on the basis of retrospective analysis of decisions [46]. Stated preference methods can be broadly classified as compositional and decompositional (see Helm et al. [47] and Weernink et al. [48]). Compositional methods involve eliciting stakeholders' preference for criteria separately. In this instance, scoring is undertaken separately from weighting. Decompositional methods involve eliciting stakeholders' preferences for whole alternatives, from which the combined weights and scores for criteria are derived simultaneously. Good practice guidelines are already available to support the implementation of decompositional methods. For instance, the ISPOR good practice guidelines on implementing DCEs [49,50].

The selection of the scoring method will depend on a number of characteristics of the decision problem:

1. Whether scoring functions or direct rating is required. Scoring functions define the score that will be attributed to all levels of performance along a criterion, and can be generated using difference or bisection approaches. Using functions makes the relationship between performance on a criterion and preference for that performance transparent. Alternatively, the performance of an alternative can be scored directly. In this case, rather than generating a function that defines the score for all levels of performance, scores are estimated for just the performance of the alternatives being evaluated.

2. The level of precision required. This is partly a function of whether the objective of the MCDA is to rank or value alternatives. Precise valuations are required for pricing decisions or designing an HTA methodology. Less precise preferences may be sufficient to inform the ranking of alternatives required by regulatory decisions or SDM. However, this will depend on how different the alternatives are—ranking alternatives with only marginal differences will require greater precision.

   The precision of scoring methods depends on whether they display interval properties. Scores have interval properties when equal increments have equal value. This is easier to achieve with approaches that generate interval or ratio scales, such as partial value functions, point allocation methods, or the coefficients generated by the DCE [51]. Approaches that adopt ordinal scales, such as analytical hierarchy process (AHP), do not necessarily display interval properties. This is also easier to achieve when scoring functions are linear. There are circumstances when this is more likely: when a criterion is a fundamental objective of value in itself (for instance, number of lives saved) or when the range being valued is very small (for instance, where cost is small compared with the decision makers' budget) [17].

3. The cognitive burden posed to stakeholders. The behavioral decision literature has identified various challenges experienced by stakeholders faced with preference elicitation questions (see Weber and Borcherding [52], Morton and Fasolo [38], and Montibeller and von Winterfeldt [27]). A number of health care MCDAs that involved patients identified this challenge as influencing the design of the MCDA study [53,54], and a potential reason for inconsistencies observed in the preference data [21,39,55]. However, MCDA studies that have surveyed patients about elicitation tasks [56,57] suggested that patients were able and willing to provide the required data.

   Cognitive burden may be less of a concern where stakeholders have experience of making the judgments required by the scoring and weighting tasks. But it is still important that the analyst support stakeholders to elicit valid scores and weights. For

Table 2 – Typology of scoring and weighting techniques used in value measurement MCDA models.

Category		Method	Scoring task	Weighting task	Examples in health care
Decompositional	Choice based	DCE/conjoint analysis	Which alternative is preferred, given the performance of each on all criteria		Baltussen et al. [102], Marsh et al. [103], Defechereux et al. [104], Mühlbacher et al. [105], Cleemput et al. [106]
		PAPRIKA	Which alternative is preferred, given the performance of each on two criteria		Hansen et al. [107], Golan and Hansen [108], Johnson et al. [109], French et al. [110]
		Best–worst scaling	Which is the worst and best alternative from three or more choices, given the performance of each on all criteria		Swancutt et al. [111], Al-Janabi et al. [112]
Compositional	Ranking	SMARTER	Not usually used for scoring	Rank order of criteria	Zuniga et al. [113]
	Direct rating	Scales	The importance of alternatives on each criterion is considered on a scale, such as a visual analogue scale (VAS)	Importance of each criterion considered separately on a scale	Goetghebeur et al. [21]
		Point allocation, e.g., SMART	Points are allocated to alternatives in proportion to their relative importance on a criterion	Allocation of points between criteria in proportion to their relative importance	Wilson et al. [114], Sussex et al. [54], Kroese et al [115], Bots and Hulshof [116], van Til et al. [117]
	Pairwise	AHP	Alternatives are compared pairwise on each criterion and their "intensity of importance" relative to each other is usually expressed on a 1–9 ratio scale	Pairwise comparison of the "intensity of importance" of criteria on a 1–9 ratio scale	Dolan et al. [71], van Til et al. [118], Hummel et al. [70]
		MACBETH	Pairwise comparison of alternatives on each criterion to assess their relative importance using sematic categories	Pairwise comparison of the "intensity of importance" using seven qualitative (semantic) categories of importance	Pinheiro et al. [119], Bana e Costa et al. [120], Oliveira et al. [121]
	Swing weighting	SMARTS	Not used for scoring	Relative importance of ranges of performance on each criteria (the "swing")	European Medicines Agency [122], Felli et al. [123]
	Scoring functions	Bisection and difference methods	The range of performance on a criterion defines the 0 and 100 points on the scoring function. The shape of the function is determined by 1) bisection: identify the performance level that is worth 50 and 2) difference: identify the score on the 0–100 scale for the midpoint on the range of performance. These steps are then repeated for the subscales to define the shape of the scoring function	Not used for weighting	Tervonen et al. [10]

AHP, analytical hierarchy process; DCE, discrete choice experiment; MACBETH, Measuring Attractiveness by a Categorical Based Evaluation TecHnique; MCDA, multiple criteria decision analysis; PAPRIKA, Potentially All Pairwise RanKings of all possible Alternatives; SMART, Simple Multi-Attribute Rating Technique; SMARTER, SMART exploiting ranks; SMARTS, SMART with Swings.

instance, eliciting committee members' preferences as inputs into an MCDA designed to support regulatory decisions may be easier than if patients' preferences are used instead, because committee members are experienced at making benefit-risk trade-offs. Equally, patients' experience of treatment outcomes may put them in a better place to understand the trade-offs being made. Further research is required to understand the cognitive burden posed by elicitation tasks, how this varies between methods and types of health care stakeholders, the impact this has on results of the MCDA, and how the analyst can mitigate this burden. In the meantime, it is important that analysts pilot elicitation tasks before they are used to collect stakeholder preferences.

5a. Report and justify the methods used for weighting

The objective of weighting is to capture stakeholders' preferences *between* criteria. That is, weights represent the "trade-offs" or "exchange rates" that bring individual criterion value scores to a common value scale. Reviews of MCDAs in health care have identified a need for more work to support the selection of appropriate weighting methods [5].

The need to consider cognitive burden on stakeholders (see step 4a) also applies when selecting weighting methods. Two further considerations are as follows:

1. Level of precision. The precision of weighting methods depends on whether they generate scaling constants—reflecting the rate at which changes in criteria compensate one another. Weights are more likely to be scaling constants when they are based on elicitation tasks that take account of the range of performance of alternatives, and that require stakeholders to trade-off changes in criterion for changes in other criteria, rather than assessment of the importance of criteria [18]. These conditions are best met by the swing weighting and decompositional approaches [51]. AHP elicits weights before ranges for performance for criteria have been set [2]. Methods that do not meet these requirements, such as direct rating, tend to produce flatter weight distributions, with criterion receiving more similar weights [58].
2. Theoretical foundations: Choice-based and swing weighting methods are based on multiattribute utility theory or multiattribute value theory [59–62]. They provide procedures to bring decision making in practice closer to the normative ideal of coherent choices. Specifically, they are based on a number of axioms that describe coherent choices, including completeness, transitivity, and independence [59]. Within utility theory-based methods, DCE differs from, for instance, swing weighting because it is based on random utility theory [63]. This acknowledges an element of randomness to observed choices due to the researchers' inability to identify all influences. Other methods diverge from the axioms of utility theory. Some direct rating approaches, such as the use of the visual analogue scale, are based on psychometric theory. AHP has a different theoretical basis [64,65], a key difference from multiattribute value theory/multiattribute utility theory being that it does not require that preferences be transitive (if x is preferred to y, and y is preferred to z, then x must be preferred to z) [12,15,66]. As a consequence, the results of AHP are subject to rank reversal—changes in the ranking of alternatives when a new alternative is introduced [2]. It is important to ensure that the theory underlying a method is consistent with decision makers' objectives. HTA is perhaps the decision where most theoretical work has been undertaken. See, for instance, the extrawelfarist foundations of cost-utility analysis [67] though a welfarist foundation has also been suggested for benefit-risk assessment [7]. It has been demonstrated that cost-utility analysis based on welfarist foundations is a special case of

multicriteria methods [68]. Further work is required on the appropriate theoretical basis for many of the health care decisions of interest in this report, and the implications for the use of MCDA.

Stakeholder heterogeneity will also impact the selection of weighting methods and how it is implemented. Existing MCDAs in health care demonstrate that preferences vary both between stakeholders types, such as between experts and patients [54,69,70], and within stakeholder groups, such as patients [39,56,57]. The authors of these studies also reflect on the implications for elicitation methods and sampling strategies; including a single stakeholder workshop may be insufficient to ensure a representative assessment, and multiple stakeholders workshops or surveys may be necessary [70]; it may be necessary to elicit preferences from patients from multiple practices [71].

6a. Report and justify the aggregation function used

The objective of aggregation is to select the appropriate function that allows scores and weights to be combined in a way that is consistent with stakeholders' preferences [72]. This step is not required for AHP, for which inputs are matrices of paired comparisons, which are analyzed using matrix algebra [72].

The most commonly applied aggregation formula in health care MCDAs is the additive model. This is the case with both composition and decompositional approaches. Additive functions are also commonly applied in instruments to estimate health-related quality of life, such as the EuroQol five-dimensional questionnaire and the six-dimensional health state short form (derived from the 36-item short form health survey) [73]. The form of an additive function is given below:

$$V_j = \sum_{i=1}^{n} S_{ij} \cdot W_i$$

where V_j is the overall value of intervention j, s_{ij} is the score for intervention j on criterion i, and w_i is the weight attached to criterion i.

Additive models have the advantage of being easy to communicate to decision makers, but impose a number of restrictions on the calculation of overall benefit, in particular the requirement that criteria be preferentially independent (see step 2a). Where this is not the case, or when an individual criterion is of primary importance or has a significant impact on overall benefit, multiplicative functions can be adopted [17,58]. Multiplicative models are also used in health-related quality-of-life instruments, such as the health utilities index [74]. The functional form of a multiplicative model varies. One example of a multiplicative model applied in an MCDA in health care is that used by Peacock et al. [72] to evaluate a South Australian community health service:

$$U = U_h[1 + W_1 D_1 \ldots W_n D_n]$$

where U is the estimate of overall value, U_h is the score for impact on individual health, $D_1 - D_n$ are scores on other criteria, and $W_1 - W_n$ are weights on other criteria. This model has the property that if individual health gain is zero, U is also zero. Another example of multiplicative MCDA models in health care is the ISafE approach used to determine what is included in the Thai essential drug list [75].

Multiplicative models are less frequently applied in practice because determining the functional form of a multiplicative model and estimating the parameters required to populate them are considered more complex than in additive models [2,76]. This has led to the use of pragmatic simplifications, and the use of additive models in which interactions between criteria are only weak [77]. However, others have argued that multiplicative models can be used in health care in a user-friendly manner [29]. How to work with stakeholders to identify the functional form of a multiplicative

model that corresponds with their preferences is, however, a topic that has been overlooked in the health MCDA literature and should be part of an ongoing research agenda.

The design of the aggregation approach should also address how to deal with heterogeneous preferences. The MCDA literature includes three types of group decision-making methods that differ in how they deal with heterogeneity ([78]; see also Phillips [79]): sharing, in which decision makers act as one decision maker, and one value for scores and weights is agreed upon; aggregating, in which individual judgments are retained and aggregated in the final outcomes, using, for instance, the mean of preferences; and comparing, in which individual judgments are retained and results for individuals compared. Compositional approaches allow each of these approaches. Decompositional approaches, such as DCE, tend to apply the averaging approach, with preference heterogeneity being reflected in either coefficient variance or the error term, unless subgroup analysis is undertaken or respondent characteristics are interacted with treatment attributes in the regression function. The choice of aggregation method is important because it can impact model outcomes [80] and will depend on the decision problem.

7a. Report sources of uncertainty

All MCDAs are subject to uncertainty, and the systematic examination and reporting of the uncertainty are hallmarks of good practice. Existing typologies of sources of uncertainties (e.g., Briggs et al. [81]) are helpful in understanding the sources of uncertainty that may impact an MCDA. The types of uncertainty that may impact the results of an MCDA should be reported, including the following:

1. Imprecise or incomplete model inputs, such as standard errors around measures of performance, or stakeholders' inability to provide precise weights or scores (stochastic and parameter uncertainty in the Briggs et al. [81] typology).
2. Variability in model inputs, such as different performance measures for subgroups of patients treated with a drug, or a divergence of opinions on weights or scores ("heterogeneity" in the Briggs et al. [81] typology).
3. Quality of evidence, such as relying on expert opinion to estimate performance measurement.
4. Structural uncertainty, such as disagreement on the weighting method or the value tree.

7b. Report and justify the uncertainty analysis

Two broad approaches to considering the impact on uncertainty are available: including uncertainty as a criterion in the MCDA and sensitivity analysis. Which of these approaches is appropriate will depend on the risk attitudes of stakeholders and the ease of capturing and communicating multiple forms of uncertainty in a single criterion.

There are several methods for understanding the impact of uncertainty on the results of an MCDA. First, a "confidence" criterion can be included in the model, reflecting the risk that the benefits captured by the other criteria will not be attained [2]. This acts as a negative penalty score that becomes more negative the greater the risk. In health care this approach has been adopted by, for instance, the Evidence and Value: Impact on Decision Making (EVIDEM) framework designed to support HTA [21]. This has the advantage of not only reflecting the impact of uncertainty in the model but also capturing stakeholders' risk attitudes via the scores and weights for this criterion. A challenge with incorporating uncertainty into MCDA as a criterion is that all the elements of the MCDA will be subject to uncertainty, and capturing all this uncertainty on a single scale may be difficult, itself requiring an assessment of the relative value of these different sources, and may obscure from stakeholders the precise sources of uncertainty. It is also worth considering whether preferences for an uncertainty

criterion would be independent of other criteria, such as effectiveness. For instance, would the preference attached to certainty increase if the effectiveness of an alternative was marginal?

Where stakeholders display little risk aversion, scoring and weighting can take the form of preferences for certain consequences [17], and it is not necessary to incorporate risk into the criteria list. In principle, HTA bodies making a large number of decisions should be risk-neutral because the chance of underestimating the value of a technology is as much as the chance of overestimating it [82]. However, some HTA bodies (such as Institute for Quality and Efficiency in Health Care (IQWiG)) have signaled a preference about certainty of outcomes and have tended to reward technologies with more certain outcomes [82]. This preference about uncertainty is supported by surveys of health care decision makers [26] and the design of MCDAs in health care [5].

A second approach, especially when stakeholders are risk-neutral, is to use one of several types of uncertainty analysis to explore how the results of the analysis will vary as a result of uncertainty (see Briggs et al. [81], Durbach and Stewart [83], Broekhuizen et al. [84], and Grouthuis-Oudshoorn et al. [85]). For instance, deterministic or probabilistic sensitivity analysis can be used to explore the impact of parameter imprecision and variability. At a minimum, deterministic analysis should be undertaken. Whether it is appropriate to also undertake a probabilistic analysis will depend on whether uncertainty in multiple parameters needs to be taken into account simultaneously, and whether dependence exists between parameters [84]. The impact of structural uncertainty can be explored by re-running the analysis using, for instance, different weighting methods.

Where preferences inputs are incomplete—for instance, weights are missing or stakeholders have provided imprecise preferences, such as ordinal ranking of criteria—inverse-preference approaches can be used to provide information on the types of preferences that would lead to the selection of particular alternatives, for instance, using stochastic multicriteria acceptability analysis [10].

8a. Report the MCDA method and findings

The results of the MCDA should be interpreted on the basis of a transparent reporting of the MCDA. The checklist (Table 1) identifies the elements of MCDA that should be reported. The results of the MCDA should be accessible to decision makers. Without contextualizing with a transparent description of scoring and weighting methods, the results of an MCDA can be difficult to interpret [15]. Communication of the inputs and outputs of an MCDA can be supported by the use of several tabular and graphical formats, a detailed survey of which can be found in the reports of the Innovative Medicines Initiative, Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (IMI PROTECT) project [86,87]. The generation of these results can be supported by the use of relevant software (see the "Resources, Skills, and Software" section).

The presentation of results should consider the decision problem. For instance, the decision problem may call for a ranking of alternatives, though this may also be supported by other outputs, such as the probability that an alternative ranks first. Other problems may require an assessment of the relative value of alternatives. For instance, benefit-cost ratios or efficiency frontiers may be used to inform resource allocation decisions (see the "Other Considerations When Designing an MCDA" section).

8b. Examine the MCDA findings

MCDA is intended to serve as a tool to help decision makers reach a decision—*their* decision, not the tool's decision (though the first task force report identified exceptions to this rule [1]). This can be facilitated by presenting the MCDA model to decision makers and

allowing them to explore the results and their sensitivity to different inputs. This is particularly useful when the MCDA yields surprising results, allowing the decision maker to explore the reasons for the discrepancy with their expectations. The examination of results can be supported by the use of relevant software (see the "Resources, Skills, and Software" section).

Other Considerations When Designing an MCDA

The Order of MCDA Steps

The order of the checklist should not be taken to imply a particular order that should be followed when implementing MCDA. First, good practice may require an iterative approach to MCDA design. For instance, the scoring exercise may reveal a lack of independence between criteria because stakeholders are unable to score changes on criterion without knowing performance on other criteria. In this case, it may be necessary to redefine criteria or adopt a different aggregation function.

Second, it may not be necessary to complete all the steps to support decision making (the notion of requisite modeling is based on a similar principle [88]). Defining the decision problem, selecting criteria, and measuring performance (what we refer to as "partial MCDA"), and not explicitly scoring and weighting criteria and calculating aggregate scores, may be sufficient. It may be unnecessary to undertake explicit scoring and weighting to support decision making if the partial MCDA reveals an alternative that performs better on all criteria, or clearcut trade-offs. When this is not the case, it is good practice to undertake the remaining steps of the MCDA.

Third, it may not be possible to undertake performance measurement and scoring before eliciting weights. In a reusable model, weights and scoring rules will need to be elicited for a plausible range of performance before the performance of alternatives is measured. The measurement and scoring of the performance of an alternative is then undertaken when the tool is applied.

Fourth, directly scoring alternatives before weighting may result in weights being influenced by knowledge of the performance of alternatives if they are not anonymized during the scoring. If anonymizing alternatives are not feasible, it may be preferable to undertaking weighting before scoring.

Dealing with Budget Constraints

Several health care decisions are subject to a budget constraint including HTA and commissioning, and some SDM require consideration of patient out-of-pocket costs. However, best practice to consider budget constraints in MCDA is still unclear, and further research should focus on this topic.

A large proportion of MCDAs in health care address budget constraints by including cost as a criterion. A recent review found that of 23 MCDAs undertaken to support health care reimbursement and coverage decisions, 10 studies included cost as a criterion [5]. This is equivalent to asking stakeholders to estimate willingness-to-pay values for the benefits. It may be feasible for an individual patient to undertake this trade-off in the context of an MCDA for SDM because they are aware of their budget constraint and the alternative uses of funds. It has, however, been argued that in other situations this approach does not adequately capture the opportunity cost of alternatives [89]. That is, stakeholders do not have the knowledge to estimate the benefits that would have to be forgone to fund an alternative. Instead, this would require the forgone alternatives to be identified and evaluated using the same MCDA framework.

It is possible to envisage this approach being applied where the alternatives that would be disinvested to fund a new alternative can be identified and evaluated. Program budgeting and marginal analysis is an illustration of this use of MCDA [72,90]. In this case, investment and disinvestment options are identified, they are evaluated using an MCDA, and ranked on the basis of the ratio of their cost to their MCDA-derived benefit. A similar logic is followed by IQWiG, which has suggested using MCDA to estimate the aggregate benefit for treatments available for a specific indication, which is then combined with cost in an efficiency frontier [91]. However, where it is not possible to identify options for disinvestment, such as with many HTAs, measuring the forgone opportunity becomes more difficult. In this case, further work would be required to estimate this opportunity cost. MCDA is not a solution to the challenge of estimating opportunity cost, and this challenge is not limited to MCDA, as illustrated by the ongoing debate about the difficulty of measuring opportunity cost in the context of cost-utility analysis [92].

Constructing a cost-benefit ratio using MCDA outputs faces several challenges. First, different scales are used to measure benefits and costs [90]. For instance, the benefit estimate generated by an MCDA may be estimated for a single recipient of an alternative, and may be restricted to a 0 to 1 or a 0 to 100 scale. If the ratio of costs and MCDA-derived benefit is used, rather than being incorporated as a criterion into an MCDA, costs are not similarly restricted, and if estimated at a system level (reflecting the number of patients who will receive an alternative) will overestimate costs in comparison with benefits. Such an analysis will be biased toward cheaper alternatives. It is important when comparing costs and benefits to ensure that they are estimated in as comparable a manner as possible; for instance, both could be estimated on a per-patient basis. Second, and a related challenge, there may be scale insensitivity in the assessment of benefit that does not also impact the assessment of cost [38]. This is particularly the case when less precise scoring and techniques are used (see step 5a). Third, it is necessary to use a ratio scale to measure benefits; otherwise, it is not possible to say that alternatives with a lower cost-benefit ratio are necessarily more efficient (see Morton [93] and step 4a for the challenges comparing the relative value of points on nonratio scales).

Resources, Skills, and Software

The successful implementation of MCDA requires the time of the analyst, experts, stakeholders, and decision makers. Different dimensions of the decision context will influence the resources available:

1. The time available to make a decision will vary between problems. For instance, HTA decisions have more time and resources available to them than do share decisions between a clinician and an individual patient.
2. The resources available to support a decision will vary between decisions and locations. More resources are likely be to be made available by higher-income countries compared with lower- and middle-income countries; national-level decision makers compared with regional-level or local-level decision makers; and to support reusable models rather than one-off decisions.

The design of the MCDA should plan for the following resource and skills requirements:

1. Analyst: The analyst needs the time, technical expertise, and appropriate software to successfully implement the chosen method. Invariably, MCDA will require a multidisciplinary team. The types of competencies required include 1) decision

analysis; 2) identifying, reviewing, and synthesizing evidence; 3) workshop facilitation; 4) survey design; 5) behavioral decision theory; and 6) statistical analysis, for instance, the use of regression models to analyze the results of DCEs.

2. Stakeholders: The success of the MCDA will rely on the commitment of stakeholders, who will have other calls on their time. A workshop may require stakeholders to be available at the same time. A survey-based method may be less demanding on stakeholders' time.

3. Experts: The multidisciplinary nature of MCDA means that the analysts' own expertise may require supplementing by expertise in the therapeutic area of interest and in the methods being used.

Many steps and recommendations outlined in the ISPOR MCDA Good Practice Guidelines Checklist can be supported by specialized decision-making software, and these are described in detail elsewhere [94,95]. In addition, most MCDA software packages are available on the Internet [96] and offer free trials. These sometimes include demonstration models and offer excellent opportunities to directly experience MCDA. The software is especially useful for problems involving relatively large numbers of alternatives and criteria, and when using weighting and scoring, and can support the generation of graphical and tabular outputs. Some of the software packages also support survey development and collection of criteria weights. Although it is important to be aware that software packages rarely allow all MCDA methods to be applied, it is recommended that the appropriate approach be determined first, and the software package selected accordingly.

Research Directions

The use of MCDA in health care is in its infancy, and so any good practice guidelines can only be considered "emerging" at this point. As a consequence, this task force article draws good-practice guidelines from the broader MCDA literature and interprets them in the light of what is known about the characteristics of health care decisions. Inevitably, then, several areas for further research are identified, including the level of precision required of an MCDA; the cognitive challenges facing different types of stakeholders and the support that can overcome these challenges; decision makers' preferences for the theoretical foundations of MCDA methods; which value functions best describe stakeholders preferences; and the best methods for incorporating uncertainty and budget constraints into an MCDA.

We would recommend that further research be undertaken in two stages. First, a productive first step would be to undertake further secondary research to address these issues separately for each type of health care decision, something that was beyond the scope of the first two task force reports. Second, unanswered research questions will likely remain, which would benefit from primary research. For instance, to date there has been very little work on the impact that MCDA has on decision making. The only evaluation of decision making with and without MCDA that we are aware of is a comparison of the use of MCDA or educational interventions in SDM [97]. Other pilot work has surveyed participants for their perception of MCDA, but not in comparison with other methods (for instance, Goetghebeur et al. [21]). Similarly, we know of only four studies that compared MCDA methods in health care [98–101].

Finally, the report focuses on value measurement approaches because other methods are rarely applied in health care. It is not clear whether this focus is appropriate. Further work should be undertaken to ensure that the conditions under which value measurement approaches are appropriate (in particular, compensatory criteria) actual hold for health care decisions.

Conclusions

The first task force report defined MCDA; considered the motivation for its use; identified the steps commonly involved in undertaking MCDA; and illustrated the diversity of approaches used in health care. This second task force report provides good practice guidance on how to select and implement appropriate MCDA techniques. A checklist is provided to guide the design and reporting of MCDAs. Although it is possible to identify good practices that should inform the use of MCDA in health care, inevitably this endeavor would benefit from further research. This task force report will support the translation of good practice guidelines into practical recommendations for how MCDA should be undertaken in different health care contexts.

Acknowledgments

R E F E R E N C E S

[1] Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, Longrenn T, Mussen F, Peacoc S, Watkins J, IJzerman M. Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health 2016;19(1):1–13.

[2] Dodgson J, Spackman M, Pearman A, Phillips L. Multi-criteria analysis: amanual. 2009. Available from: http://www.communities.gov.uk/documents/corporate/pdf/1132618.pdf. [Accessed September 13, 2015].

[3] Phillips LD, Bana e Costa CA. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. Ann Oper Res 2007;154:51–68.

[4] Bana e Costa CA, Lourenço JC, Oliveria MD, Bana e Costa JC. A socio-technical approach for group decision support in public strategic planning: the Pernambuco PPA case. Group Decis Negot 2014;23:5–29.

[5] 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.

[6] Dolan J. Multi-criteria clinical decision support: a primer on the use of multi-criteria decision making methods to promote evidence-based patient centered healthcare. Patient 2010;3:229–48.

[7] Hauber AB, Fairchild AO, Johnson FR. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy 2013;11:319–29.

[8] Ivlev I, Kneppo P, Bartak M. Multicriteria decision analysis: a multifaceted approach to medical equipment management. Technol Econ Dev Econ 2014;20:576–89.

[9] van Til J, Groothuis-Oudshoorn CGM, Lieferink M, et al. Does technique matter: a pilot study exploring weighting techniques for a

multi-criteria decision support framework. Cost Eff Resour Alloc 2014;12:22.

[10] Tervonen T, Naci H, van Valkenhoef G, et al. Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention. Med Decis Making 2015;35:859–71.

[11] Belton V, Stewart TJ. Multiple Criteria Decision Analysis: An Integrated Approach. Kluwer Academic Publishers, MA, 2002.

[12] Guitouni A, Martel J-M. Tentative guidelines to help choosing an appropriate MCDA method. Eur J Oper Res 1998;109:501–21.

[13] Velasquezl M, Hester PT. An analysis of multi-criteria decision making methods. Int J Oper Res 2013;10:56–66.

[14] De Montis A, De Toro P, Droste-Franke B, et al. Criteria for quality assessment of MCDA methods. Presented at: 3rd Biennial Conference of the European Society for Ecological Economics. Vienna, Austria, May 3-6, 2000.

[15] De Montis A, De Toro P, Droste-Franke B, et al. Assessing the quality of different MCDA methods. In: Getzner M, Spash C, Stagl S,eds. Alternatives for Environmental Evaluation. Abingdon, Oxon, UK: Routledge, 2005.

[16] Getzner M, Splash CL, Stagl S. Alternatives for environmental valuation. New York: Routledge, 2005.

[17] Keeney RL, von Winterfeldt D. Practical value models: published articles and papers. 2009. Paper 36. Available from: http://research. create.usc.edu/published_papers/36. [Accessed August 14, 2015].

[18] Keeney RL. Common mistakes in making value trade-offs. Oper Res 2002;50:935–45.

[19] Olson DL, Mechitov AI, Moshkovich H. Comparison of MCDA paradigms. In: Advances in Decision Analysis. Springer, Netherlands, 1999.

[20] Phillips LD. Best practice for MCDA in healthcare. In: Marsh K, Thokala P, Goetghebeur M, Baltussen B, eds., Healthcare Decisions Supported by Multi-criteria Decision Analysis. Springer. In Press.

[21] 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;32:376–88.

[22] Franco LA, Montibeller G. Problem structuring for multicriteria decision analysis interventions. In: Cochran JJ, ed., Wiley Encyclopedia of Operations Research and Management Science. John Wiley and Sons, 2010.

[23] National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. London, UK: National Institute for Health and Care Excellence, 2013.

[24] National Institute for Health and Care Excellence. Available from: https://www.nice.org.uk/media/default/About/what-we-do/Research-and-development/Social-Value-Judgements-principles-for-the-development-of-NICE-guidance.pdf. [Accessed September 12, 2015].

[25] Youngkong S, Teerawattananon Y, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. Health Res Policy Syst 2012;10:6.

[26] Tanios N, Wagner M, Tony M, et al. Which criteria are considered in healthcare decisions? Insights from an international survey of policy and clinical decision makers. Int J Technol Assess Health Care 2013;29:456–65.

[27] Montibeller G, von Winterfeldt D. Cognitive and motivational biases in decision and risk analysis. Risk Analysis 2015;35:1230–51.

[28] Mullen PM. Quantifying priorities in healthcare: transparency or illusion? Health Serv Manag Res 2004;17:47–58.

[29] Morton A, Lauer JA. Incorporating social values additional to health in health benefits packages. In: Glassman A, Giedion U, Smith P, eds., The How-to of Health Benefits: Options and Experiences on the Path to Universal Health Coverage in Low- and Middle-Income Countries. Washington, DC: Center for Global Development. In Press.

[30] Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health 2009;12:S10–5.

[31] Morton A. Aversion to health inequalities in healthcare prioritization: a multiplicative mathematical programming perspective. J Health Econ 2014;36:164–73.

[32] Bana e Costa CA, Lourenço JC, Chagas MP, Bana e Costa JC. Development of reusable bid evaluation models for the Portuguese Electric Transmission Company. Dec Analysis 2008;5:22–42.

[33] Mt-Isa S, Wang N, Hallgreen CE, et al. Review of Methodologies for Benefit and Risk Assessment of Medication. London, UK: PROTECT Consortium, 2014.

[34] Berkeley D, Humphreys P. Structuring decision problems and the "bias heuristic.". Acta Psychol 1982;50:201–52.

[35] Stillwell WG, von Winterfeldt D. Comparing hierarchical and nonhierarchical weighting methods for eliciting multiattribute value models. Manag Sci 1987;33:442–50.

[36] Winterfeldt, von D, Fasolo B. Structuring decision problems: a case study and reflections for practitioners. Eur J Oper Res 2009;199:857–66.

[37] Hughes D, Waddingham Ed AJ, Mt-Isa Shahrul, et al. on behalf of IMI-PROTECT Work Package 5. Recommendations for the Methodology and Visualization Techniques to be Used in the Assessment of Benefit and Risk of Medicines. IMI-PROTECT Benefit-Risk Group. London, UK, 2013.

[38] Morton A, Fasolo B. Behavioural decision theory for multi-criteria decision analysis: a guided tour. J Oper Res Soc 2009;60:268–75.

[39] Hummel JM, Steuten LG, Groothuis-Oudshoorn CJ, et al. Preferences for colorectal cancer screening techniques and intention to attend: a multi-criteria decision analysis. Appl Health Econ Health Policy 2013;11:499–507.

[40] De Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012;21:145–72.

[41] Keeney RL. Value-Focused Thinking: A Path to Creative Decision Making. Harvard University Press 1992, Cambridge, MA.

[42] 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 Serv Res 2011;11:329.

[43] Busse R, Orvain J, Velasco M, Perleth M. Best practice in undertaking and reporting health technology assessments. Working group 4 report. Int J Technol Assess Health Care 2002;18:361–422.

[44] Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.

[45] European Medicines Agency. Benefit-Risk Methodology Project. Work Package 4 Report: Benefit-Risk Tools and Processes. 2012. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf.

[46] Dakin H, Devlin N, Feng Y, et al. The influence of cost-effectiveness and other factors on nice decisions. Health Econ 2015;24(10):1256–71.

[47] Helm R, Scholl A, Manthey L, Steiner M. Measuring customer preferences in new product development: comparing compositional and decompositional methods. IJPD 2004;1:12.

[48] Weernink M, Janus S, van Til JA, et al. A systematic review to identify the use of preference elicitation methods in healthcare decision making. Pharmaceut Med 2014;28:175–85.

[49] Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011;14:403–13.

[50] 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 2013;16:3–13.

[51] Marsh K, Thokala P, Mühlbacher A, Lanitis T. Incorporating preferences and priorities into MCDA: selecting an appropriate scoring and weighting technique. In: Marsh K, Thokala P, Goetghebeur M, Baltussen B, eds., Healthcare Decisions Supported by Multi-criteria Decision Analysis. Springer. In Press.

[52] Weber M, Borcherding K. Behavioural influences on weight judgments in multiattribute decision making. Eur J Oper Res 1993;67:1–12.

[53] Airoldi M, Morton A, Smith J, Bevan G. Working Paper No. 7. Healthcare prioritisation at the local level: a socio-technical approach. 2011. LSE Department of Management, Priority Setting for Population Health Working paper series. Working paper no. 7. Available from http://www.lse.ac.uk/management/documents/research/ref2014-impact-case-studies/WP7_-_Healthcare_prioritisation_at_the_local_level_A_socio-technical_approach.pdf.

[54] Sussex J, Rollet P, Garau M, et al. A pilot study of multicriteria decision analysis for valuing orphan medicines. Value Health 2013;16:1163–9.

[55] Youngkong S, Baltussen R, Tantivess S, et al. Multicriteria decision analysis for including health interventions in the universal health coverage benefit package in Thailand. Value Health 2012;15:961–70.

[56] Dolan JG. Patient priorities in colorectal cancer screening decisions. Health Expect 2005;8:334–44.

[57] Dolan JG, Boohaker E, Allison J, Imperiale TF. Patients' preferences and priorities regarding colorectal cancer screening. Med Decis Making 2013;33:59–70.

[58] Von Winterfeldt D, Edwards W. Decision Analysis and Behavioural Research. New York: Cambridge University Press, 1986.

[59] von Neuman J, Morgenstern O. Theory of Games and Economic Behavior (2nd ed.). Princeton, NJ: Princeton University Press, 1947.

[60] Krantz DH, Luce RD, Suppes P, Tversky A. Foundations of Measurement, volume I. New York: Academic Press, 1971.

[61] Keeney RL, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: Wiley, 1976: Reprinted in 1993 by Cambridge University Press.

[62] Dyer JS, Sarin R. Measurable multiattribute value functions. Oper Res 1979; 22:810–22.

[63] Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, eds. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht, The Netherlands: Springer, 2008.

[64] Saaty TL. Axiomatic foundation of the analytic hierarchy process. J Manag Sci 1986;32:841–55.

[65] Saaty TL. Decision making with the analytic hierarchy process. Int J Serv Sci 2008:1.

[66] Vargas LG. Priority theory and utility theory. Mathematical Modelling 1987;9:381–5.

[67] Brouwer WB, Culyer AJ, van Exel NJ, Rutten FF. Welfarism and extra-welfarism. J Health Econ 2008;27:325–38.

[68] Postmus D, Tervonen T, van Valkenhoef G, et al. A multi-criteria decision analysis perspective on the health economic evaluation of medical interventions. Eur J Health Econ 2014;15:709–16.

[69] Hummel JM, Snoek GJ, van Til JA, et al. A multicriteria decision analysis of augmentative treatment of upper limbs in persons with tetraplegia. J Rehabil Res Dev 2005;42:635–44.

[70] Hummel MJ, Volz F, van Manen JG, et al. Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment. Patient 2012;5:225–37.

[71] Dolan, et al. Patient priorities in colorectal cancer screening. Decisions Health Expectations 2005;8:334–44.

[72] Peacock SJ, Richardson JRJ, Carter R, Edwards D. Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). Social Science & Medicine 2007;64:897–910.

[73] Furlong W, Barr RD, Feeny D, Yandow S. Patient-focused measures of functional health status and health-related quality of life in pediatric orthopedics: a case study in measurement selection. Health Qual Life Outcomes 2005;3:3.

[74] Feeny D, Torrance G, Furlong W. Health utilities index. In: Spilker B,ed. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia: Lippencott Raven Publishers, 1996.

[75] Chongtrakul P, Sumpradit N, Yoongthong W. ISafE and the evidence-based approach for essential medicines selection in Thailand. Essential Drugs Monit 2005;34:18–9.

[76] Keeney R, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Trade-offs. Cambridge University Press, Cambridge, UK, 1993.

[77] Stewart T. Simplified approaches to multicriteria decision making under uncertainty. J Multicriteria Decis Analysis 1995;4:246–58.

[78] Belton V, Pictet J. A framework for group decision using a MCDA model: sharing, aggregating or comparing individual information. J Dec Syst 1997;6:283–303.

[79] Phillips LD. Group elicitation of probability distributions: are many heads better than one? In: Shanteau J, Mellors B, Schum D,eds. Decision Science and Technology: Reflections on the Contributions of Ward Edwards. Norwell, MA: Kluwer Academic Publishers, 1999. p. 313–30.

[80] Zhou P, Ang BW. Comparing MCDA aggregation methods in constructing composite indications using the Shannon-Spearman measure. Soc Indicat Res 2009;94:83–96.

[81] Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modelling Good Research Practices Task Force Working Group-6. Med Decis Making 2012;32:722–32.

[82] Towse A, Barnsley P. Approaches to identifying, measuring and aggregating elements of value. Int J Technol Assess Healthcare 2013;29:360–4.

[83] Durbach IN, Stewart TJ. Modelling uncertainty in multi-criteria decision analysis. Eur J Oper Res 2012;223:1–14.

[84] Broekhuizen H, Groothuis-Oudshoorn CG, van Til JA, et al. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Pharmacoeconomics 2015;33:445–55.

[85] Grouthuis-Oudshoorn CGM, Broekhuizen H, van Til J. Dealing with uncertainty in the analysis and reporting of MCDA. In: Marsh K, Thokala P, Goetghebeur M, Baltussen B, eds., Healthcare Decisions Supported by Multi-Criteria Decision Analysis. Springer. In Press.

[86] Mt-Isa S, Peters R, Phillips LD, et al. Review of Visualization Methods for the Representation of Benefit-Risk Assessment of Medication: Stage 1 of 2. London, UK: PROTECT Consortium, 2013.

[87] Mt-Isa S, Hallgreen CE, Asiimwe A, et al. Review of Visualization Methods for the Representation of Benefit-Risk Assessment of Medication: Stage 2 of 2. London, UK: PROTECT Consortium, 2013.

[88] Phillips LD. A theory of requisite decision models. Acta Psychol 1984;56:29–48.

[89] Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients. Health Econ 2015;24:1–7.

[90] Wilson ED, Peacock SJ, Ruta D. Priority setting in practice: what is the best way to compare costs and benefits? Health Econ 2009;18:467–78.

[91] IQWiG. General methods. 2015. Available from: https://www.iqwig.de/en/methods/methods-paper.3020.html. [Accessed September 14, 2015].

[92] Karlsberg Schaffer S, Sussex J, Devlin N, Walker A. Local health care expenditure plans and their opportunity costs. Health Policy 2015;119:1237–44.

[93] Morton A. Aversion to health inequalities in healthcare prioritisation: A multicriteria optimisation perspective. J Health Econ 2014;36C:164–173.

[94] Weistroffer HR, Li Y. Multiple criteria decision analysis software. In: Ehrgott M, Rui Figueira J, Greco S, eds., Multiple Criteria Decision Analysis: State of the Art Surveys (2nd ed.). Springer. In Press.

[95] Ishizaka A, Nemery P. Multi-criteria Decision Analysis: Methods and Software. Wiley, Chichester, UK, 2013.

[96] Available from: http://www.mcdmsociety.org/soft.html www.cs.put.poznan.pl/ewgmcda/index.php/software https://en.wikipedia.org/wiki/Decision_making_software.

[97] Dolan JG, Frisina S. Randomized controlled trial of a patient decision aid for colorectal cancer screening. Med Decis Making 2002;22:125–39.

[98] IJzerman MJ, van Til1 JA, Snoek GJ. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. Patient 2008;1:265–72.

[99] IJzerman MJ, van Til1 JA, Bridges JFP. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. Patient 2012;5:45–56.

[100] Van Wijk BLG, Klungel OH, Heerdink ER, de Boer A. A Comparison of Two Multiple-Characteristic Decision-Making Models for the Comparison of Antihypertensive Drug Classes Simple Additive Weighting (SAW) and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS). Am J Cardiovasc Drugs 2006;6(4):251–8.

[101] van Til J, Groothuis-Oudshoorn C, Lieferink M, Dolan J, Goetghebeur M. Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework. Cost Effectiveness and Resource Allocation 2014;12:22.

[102] Baltussen R, Ten Asbroek AHA, Koolman X, et al. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? Health Policy Plan 2007;22:178–85.

[103] Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. J Public Health (Oxf) 2013;35:460–6.

[104] Defechereux T, Paolucci F, Mirelman A, et al. Health care priority setting in Norway: a multicriteria decision analysis. BMC. Health Serv Res 2012;12:39.

[105] Mühlbacher A, Bridges J, Bethge S, et al. Choice-based Conjoint Analysis—Pilot Project to Identify, Weight, and Prioritize Multiple Attributes in the Indication "Hepatitis C". IQWiG Report 2013.

[106] Cleemput I, Devriese S, Kohn L, et al. Incorporating Societal Preferences in Reimbursement Decisions – Relative Importance of Decision Criteria According to Belgian Citizens (KCE Reports 234. D/2014/10.273/91). Brussels, Belgium: Belgian Health Care Knowledge Centre (KCE), Health Services Research, 2014.

[107] Hansen P, Hendry A, Naden R, et al. A new process for creating points systems for prioritizing patients for elective health services. Clin Gov 2012;17:200–9.

[108] Golan O, Hansen P. Which health technologies should be funded? A prioritization framework based explicitly on value for money. Isr J Health Policy Res 2012;1:44.

[109] Johnson S, Naden R, Fransen J, et al. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. J Clin Epidemiol 2014;67:706–14.

[110] French S, Bennell K, Nicolson P, et al. What do people with knee or hip osteoarthritis need to know? An international consensus list of essential statements for osteoarthritis. Arthritis Care Res 2014;67:809–16.

[111] Swancutt DR, Greenfield SM, Wilson S. Women's colposcopy experience and preferences: a mixed methods study. BMC Womens Health 2008;8:2.

[112] Al-Janabi H, Flynn TN. Coast J. Estimation of a preference-based carer experience scale. Med Decis Making 2011;31:458–68.

[113] Zuniga MA, Carrillo-Zuniga G, Seol YH, Fos PJ. Multi-criteria assessment of county public health capability disparities. J Health Hum Serv Admin 2009;32:3.

[114] Wilson E, Sussex J, Macleod C, Fordham R. Prioritizing health technologies in a primary care trust. J Health Serv Res Policy 2007;12:80–5.

[115] Kroese M, Burton H, Whittaker J, et al. A framework for the prioritization of investment in the provision of genetic tests. Public Health Genomics 2010;13:538–43.

[116] Bots PWG, Hulshof JAM. Designing multi-criteria decision analysis processes for priority setting in health policy. J Multi-criteria Dec Analysis 2000;956:75.

[117] van Til JA, Dolan JG, Stiggelbout AM, et al. The use of multi-criteria decision analysis weight elicitation techniques in patients with mild cognitive impairment: a pilot study. Patient 2008;1:127–35.

[118] van Til JA, Renzenbrink GJ, Dolan JG, Ijzerman MJ. The use of the analytic hierarchy process to aid decision making in acquired equinovarus deformity. Arch Phys Med Rehabil 2008;89:457–62.

[119] Pinheiro PR, de Castro AKA, Pinheiro MDC. A multicriteria model applied in the diagnosis of Alzheimer's disease: a Bayesian network. Presented at: 11th IEEE International Conference on Computer Science and Engineering. July 16-18, 2008. Sao Paolo, Brazil.

[120] Bana e Costa CA, Carnero MC, Oliveira MD. A multi-criteria model for auditing a predictive maintenance programme. Eur J Oper Res 2012;217:381–93.

[121] Oliveira MD, Rodrigues T, Bana e Costa CA, Sá AB. Prioritizing health care interventions: a multicriteria resource allocation model to inform the choice of community care programmes. In: Tanfani E, Testi A, eds., Advanced Decision Making Methods Applied to Health Care. Springer, Varlag, Italy, 2012:141–54.

[122] European Medicines Agency. Benefit-Risk Methodology Project. Work Package 3 Report: Field Tests. 2011. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/09/WC500112088.pdf. [Accessed September 14, 2015].

[123] Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. Med Decis Making 2009;29:104–15.