Chapter 5
Dealing with Uncertainty in the Analysis and Reporting of MCDA

Catharina G.M. Groothuis-Oudshoorn, Henk Broekhuizen, and Janine van Til

Abstract The aim of this chapter is to provide guidance regarding the various types and sources of uncertainty that influence the outcome of a multi-criteria decision analysis (MCDA) model. For each MCDA step, i.e., structuring, scoring, weighting, and aggregating, we will describe sources of uncertainty and point to methods to deal with these uncertainties. Also the use of sensitivity analyses and the relevance of qualifying and quantifying uncertainty in MCDA will be discussed. The consideration of uncertainty is a difficult but important balancing act between capturing the complex uncertainties of the decision and keeping the MCDA comprehensible for decision makers.

5.1 Introduction

Multi-criteria decision analysis (MCDA) is no exact science. The output or outcome of any decision analysis depends on assumptions and decisions made while building the model and populating that model with criteria weights and performance scores. This is often referred to with the general term "uncertainty." Uncertainty can be regarded as the lack of complete knowledge or certainty about what the model should look like and what the correct inputs are (French 1995). There are many types and sources of uncertainty that influence the outcome of the MCDA model in different ways, each of which deserves specific attention while interpreting the results of an MCDA.

This chapter first describes the different types of uncertainty. Second, it will give an overview of how the different types of uncertainty play a role in the stages of an MCDA (Chapter 4). Uncertainties in the structuring, scoring, weighting and aggregating stages are reported separately, by discussing sources of uncertainty, (appropriate)

C.G.M. Groothuis-Oudshoorn (✉) • H. Broekhuizen • J. van Til
Department Health Technology and Services Research, MIRA institute, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
e-mail: c.g.m.oudshoorn@utwente.nl

K. Marsh et al. (eds.), *Multi-Criteria Decision Analysis to Support Healthcare Decisions*, DOI 10.1007/978-3-319-47540-0_5

reporting of uncertainty, and methods to study the influence of uncertainty on the outputs of the model. Finally, the use of sensitivity analyses and the relevance of qualifying and quantifying uncertainty in MCDA are discussed. Throughout the chapter we will point readers to further readings on the different topics.

Briggs and colleagues distinguished and defined four types of uncertainty in decision analytic modeling, namely, stochastic uncertainty, parameter uncertainty, heterogeneity, and structural uncertainty (Briggs et al. 2012). We use the example of a body weight scale to illustrate these different types of uncertainty. Stochastic uncertainty is the random, unexplained variability between different measurements of the weight of one person on a single weight scale of the same type and brand that occurs as a result of randomness, like the flipping of a coin or variation in the measurements of the weight of a single person if they are measured multiple times on the same device. Parameter uncertainty refers to the variability in the estimation of a parameter of interest as a result of different interpretation of the same measurement scale, for instance, the different readings of an analog weight scale by the same person on different days or by different persons (which cannot be attributed to actual differences in weight). The distinction between stochastic and parameter uncertainty is analogous to the difference between the standard deviation, a measure of variability of individuals in a population, and the standard error, i.e., a measure of precision of an estimated quantity. Like the standard error, parameter uncertainty can usually be reduced or eliminated by increasing the number of measurements. However, like the standard deviation, stochastic uncertainty cannot be eliminated but can only be better characterized, for instance, by describing the density of the random variation or the cumulative distribution. Heterogeneity is the between-person variability that can be explained by the persons' characteristics, e.g., for weight estimates this is the difference in weight between persons as a result of their differences in body composition. Structural uncertainty refers to the notion that the output of any model is conditional on its structural assumptions with regard to the best way to reach the goal itself, for example whether it is preferred to measure weight on an analog or a digital weighting scale.

As described in the previous chapters, four stages can be identified in an MCDA: structuring, i.e., establishing the decision context and building the model, weighting, scoring, and aggregating (recommendation and sensitivity analysis) (see Fig. 5.1). In each stage of an MCDA, the different types of uncertainty can be identified.

5.1.1 Problem Structuring

MCDA is mostly used in a group decision-making setting. Belton and Pictet distinguish three types of group decision-making models that can be employed during meetings in which judgments are elicited from decision makers: sharing, aggregating, and the comparing of judgments (Belton and Pictet 1997). In *sharing*

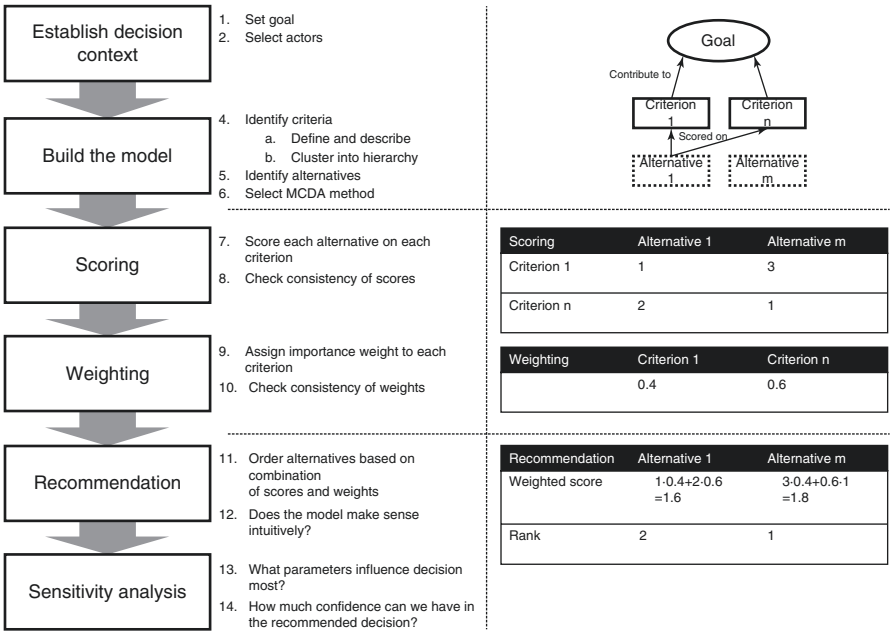

Fig. 5.1 Overview of the value-based MCDA decision-making process (*left*) and a simple numeric example (*right*)

of judgments, decision makers act as one decision maker for the purposes of the MCDA model. This implies that, even though initially there may be disagreement in the group about the judgments given, one value for each weight and performance score has to be agreed upon by the group at each stage, and only that judgment is used in the analysis. In contrast, in *aggregating judgments*, the individual judgments of each decision maker are retained throughout the decision-making process, and they are aggregated over decision makers in the final outcome, for example, by taking the mean of all individual judgments. In *comparing judgments*, the weights and performance scores of the decision makers are again retained throughout the decision analysis, and the individual judgments are actively compared during the final discussion to obtain insight into differences in opinion that may exist in between group members.

A similar but slightly different distinction in the way to handle differences in judgments is that between *statistical* aggregation and *behavioral* aggregation of judgments. Statistical aggregation is similar to Belton and Pictet's definition of aggregating: the group's individual judgments are combined into a mean judgment with a measure of variance to capture the differences between the decision makers. Behavioral aggregation is similar to Belton and Pictet's *sharing* of judgments: the group's single judgment is arrived at through a structured group process where the group can "share their knowledge and allow persuasive arguments to change their views" and therefore to revise their judgments (Phillips 1999).

The decision on which group decision model to use is a form of structural uncertainty, and moreover it influences the extent to which other types of uncertainty can be made explicit in later phases of the analysis. For instance, if at the first stage it is decided to use a sharing model where the weights and scores are set to single numbers despite possibly different judgments, the uncertainty around mean weight estimates (parameter uncertainty) and any differences between decisions makers (heterogeneity) cannot be studied. To clearly illustrate uncertainty in this chapter, we will use the statistical aggregating approach. We will illustrate the different stages of MCDA and the sources of uncertainty with a simplified case of a group decision in the medical context throughout this chapter (Text Box 5.1).

Text Box 5.1: Case Description
Six urologists within a private practice have a discussion on reducing unwanted practice variation in the choice of the first-line treatment for stage I prostate cancer patients in their practice. At present, they prescribe four alternative types of treatment: active surveillance, radical prostatectomy, external beam radiation therapy, and brachytherapy.

5.1.2 Uncertainty in Problem Structuring

The choice of criteria in an MCDA is a source of structural uncertainty. Criteria have to reflect the different points of view of the actors in the decision and enable comparisons of between the alternatives (Bouyssou 1990). To reduce structural uncertainty about whether all relevant criteria are included in the analysis, it is advised to combine the top-down and bottom-up approach to developing a set of criteria. Top-down approaches are where actors first agree on the relevance of particular consequences and then come up with examples of those often result in hierarchical value trees. Alternatively, bottom-up approaches often start with drawing up extensive lists of criteria from different sets, which can later be structured into hierarchies if desired. In the problem structuring stage, the value tree has to be determined and the final set of criteria has to be determined. Structural uncertainty about the shape of the value tree and the number and type of criteria to include can be made explicit by making detailed notes of all decisions made in this step and by including as many actors as needed to come up with a broadly supported value tree.

The goal of problem structuring is to come up with a clear, logical, and shared point of view of what decision criteria and decision structure best reflect the decision at hand and help the decision makers to achieve their objective. The final list of criteria should be as simple as possible, yet capture the complexity of the decision. There are no guidelines on what is the optimal number of criteria and/or decision structure. In some cases the type of MCDA or the cognitive limitations of the stakeholders put a limit on the number of criteria or favor a certain value tree structure. When in doubt about including a criterion, it is always wise to include it in the analysis, as some MCDA methods allow a criterion to be dropped in a later

stage. The alternative of adding a criterion in a later stage is much more bothersome. Problem structuring is a skill that is acquired through experience. Flow charts, fish-bone diagrams, pro/cons lists, and quantitative techniques, such as the nominal group technique, can help groups come up with an adequate set of criteria and an adequate problem structure (Taner et al. 2007). However, it is known that the choice of criteria, MCDA method, and weight elicitation influences the outcome of the model. Besides extensive argumentation and good documentation of the way in which the decision problem was reflected in the choice of criteria, shape of the value tree, and choice of the MCDA method, the only way to explicitly study structural uncertainty is by testing the influence of the different options (i.e., different criteria sets, value trees, and MCDA techniques) on the outcome of the analysis. There are multiple examples of such tests in literature (van Til et al. 2014; IJzerman et al. 2012a, b).

Text Box 5.2: MCDA Model and Clinical Evidence

In the example, there are many criteria that potentially influence the choice of treatment in prostate cancer. The effectiveness of the treatment in prolonging life after diagnosis; the side effects of treatment, such as bowel problems, bladder problems, erection problems, and tiredness; and the process characteristics of treatment such as costs, duration, and frequency of follow-up needed all could influence treatment preference. For illustrative purposes, we limit the example to the four criteria mentioned in Table 5.1 and choose a simple multi-attribute rating technique (SMART) to demonstrate the different types of uncertainty in weighting, scoring, and sensitivity analysis.

SMART is a simple value-based MCDA method based on a linear additive value function. In our example, the model is $V_i = \sum_{k=1}^{4} w_k x_{ik}$, where V_i is the overall value of treatment i, w_k is the weight of the kth criterion as weighted using swing weighting, and x_{ik} is the performance of treatment i on k. The hypothetical clinical evidence for the example is given in Table 5.1.

Table 5.1 Bowel problems, incontinence, and erectile dysfunction as measured as probabilities of the event occuring in five years after treatment

	Active surveillance	Surgical removal	External beam radiation therapy	Brachytherapy
Sample size	1000	800	200	800
Survival (years)	10 [9.4–10.6]	15 [14.0–16.1]	12 [10.3–13.7]	12 [11.0–13.1]
Bowel problems	0 %	0 %	15 %	0 %
Incontinence	0 %	10 %	1 %	0.5 %
Erectile dysfunction	5 %	75 %	45 %	24 %

Based on Cooperberg et al. (2012), Hayes et al. (2013)

5.2 Uncertainty in Scoring

As described in Chapter 4, during the scoring stage either the available clinical evidence or expert judgment is used to judge the performance of the alternatives on the criteria, by transforming clinical performance (which may be measured on a variety of scales) to a common value scale. Both the clinical evidence and the expert judgments are possible sources of parameter and stochastic uncertainty, as well as heterogeneity (Durbach and Stewart 2012).

5.2.1 Performance Estimates

Preferably, performance of the alternatives on the different criteria is based on clinical data (including patient registries, cost databases, etc.). In our example, the average survival for the four treatments could be drawn from scientific literature. When clinical evidence is used as input in an MCDA model, often only the point estimates are used. However, the parameter and stochastic uncertainty surrounding these estimates of performance measurements can be used to explicitly model uncertainty in the MCDA. Parameter uncertainty in the performance estimates refers to the variability in the estimation of the outcome (for instance survival) as a result of the sampling (error). The standard errors or, equivalently, the confidence bounds of the point estimates obtained from clinical trial data can be used to represent the extent of the parameter uncertainty. Stochastic uncertainty, i.e., the unexplained variability in the clinical evidence, can be made visible by presenting the standard deviation or the range of the outcomes in the patient sample.

To demonstrate any heterogeneity in the clinical evidence in the model, one can calculate averages and standard deviations of outcome for different subgroups of patients. When clinical data is lacking, the performance estimates have to be based on expert judgments. Different expert elicitation techniques are available to do so (O'Hagan et al. 2006; Bojke et al. 2010; Bojke and Soares 2014). This introduces structural uncertainty to the model due to the differences in techniques. If clinical judgments are to replace clinical evidence, rather than just asking for point estimates of performance, experts should be asked to give distributions and/or confidence bounds for their estimates, if possible linked to patient characteristics. This enables analysts to take parameter uncertainty and heterogeneity into account.

5.2.2 From Performance to Value

In valuing performance, the performance of the alternatives on the natural scale (e.g., survival in years) is transformed to a score which represents the value of that performance on a scale ranging from zero (no value) to one (maximum value). One can determine the relative value of the performance estimates for the alternatives, or

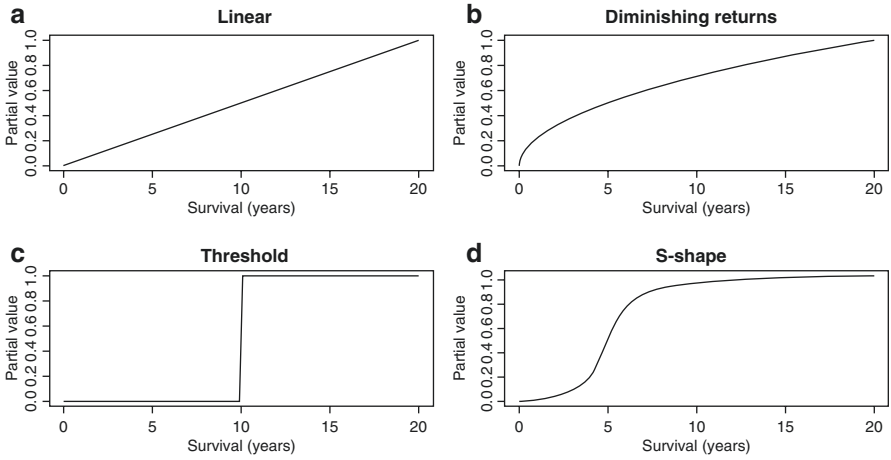

Fig. 5.2 (**a**) a partial value function where function scales linearly with increased survival. (**b**) a partial value function where there are diminishing returns with respect to survival, i.e. an increase from 0 to 5 years of survival is considered more valuable than an increase from 5 to 10 years of survival. (**c**) a partial function with a threshold. Here, all increases in survival less than 10 years are not considered valuable, but increases in survival of more than 10 years are considered valuable. (**d**) S-shaped partial value function. This can be seen as a smoothed version of the threshold function

one can map the performance to estimate value of all intermediate performances with the use of partial value functions. This can either be done "locally," meaning that the best and worst performance judgments of the alternatives on the criteria (as identified by experts) are used as the upper and lower bounds of the value function, or "globally," meaning upper and lower bounds are based on estimates of worst and best possible outcomes, irrespective of the performance of the included alternatives. For example, although a diagnostic test with sensitivity of 100 % is highly unlikely in clinical practice, 100 % sensitivity can be used as a theoretical best possible outcome.

One source of structural uncertainty in the valuation stage is the shape of the value function. Most commonly, a linear function is assumed (Fig. 5.2a). This is a simple function that linearly scales all performance values between the worst level W (partial value of zero) and the best level B (partial partial value of one):

$$v(x) = \begin{cases} 1, & x \geq B \\ \dfrac{x - W}{B - W}, & W < x < B \\ 0, & x \leq W \end{cases} \tag{5.1}$$

In such a linear value function, it is assumed that an increase in performance results in an equal increase in value independently from where on the performance scale this increase occurs. So for survival, it would mean that the value of increasing survival from 0 to 5 years is the same as the value of increasing the survival from 15 to 20 years. Confidence bounds for the value of performances can be obtained by

applying the partial value function on the confidence bounds of the performances (see Table 5.2 for the example).

However, many other forms of the value function exist. For example, there can be diminishing returns to prolonging life, and for some people, living after a certain age might decrease the incremental value of this outcome (Fig. 5.2b). For other outcomes, a value function with a threshold that may vary between decision makers may be more appropriate where performance (outcome) switches from no value to maximum value (Fig. 5.2c), or there can be a combination of a linear function and diminishing returns: the S-shaped value function (Fig. 5.2d). It is common to agree upon the shape of the functional form on a group level.

If one wants to deviate from the linear function, additional inputs are needed from decision makers to determine the particular shape of the value function. For example, in the bisection method, the decision maker is asked to define the point on the attribute scale which is halfway in value terms between the two endpoints. From this a two-piece linear value function can be constructed (Belton and Stewart 2002). This process can be repeated multiple times until the decision makers are indifferent between further bisections. In MACBETH, a value function for a particular criterion is constructed from the pairwise comparisons of the performance of alternatives on that criterion using linear programming (Costa et al. 2012).

As value judgments may differ between decision makers, the final construction of a value function can consequently be based on averages of these judgments or based on another central measure (median, mode). By calculating a standard error and confidence intervals along with the average value, a measure for parameter uncertainty can be obtained. Stochastic uncertainty in the value judgments can be quantified by calculating a standard deviation of the value judgments.

Heterogeneity refers to possible differences in the value function or value of outcomes between different groups of decision makers that may be explained with their backgrounds. For instance, thresholds for survival might be different in urologists that mostly see older patients, compared to urologists that see relatively younger patients in their daily work. By constructing value functions for the different groups, one can see whether heterogeneity is present.

Summarizing, all types of uncertainty influence the values from the scoring step. Uncertainty (parameter, stochastic, or heterogeneity) in the evidence implies also uncertainty in the value function. The uncertainty in evidence and the uncertainty due to differences in value judgments can be quantified by calculating standard errors, standard deviations, or confidence bounds.

5.3 Uncertainty in Weighting

All inputs in the weighting stage are given by stakeholders (decision makers, patients, physicians, general public, etc.), which are therefore the main source of uncertainty in this stage. In the weights there can be structural uncertainty, parameter uncertainty, stochastic uncertainty, and heterogeneity.

Text Box 5.3: Calculating Partial Values
Assuming a linear value function in the urologists' case, the point estimates with confidence bounds of performance are transformed with Eq. 5.1 to the value estimates with confidence bounds (Table 5.2)

Table 5.2 Partial values in the urologists' case, with 95 % confidence intervals for the partial values for survival based on the confidence bounds reported in the clinical trial reports

	Active surveillance	Surgical removal	External beam radiation therapy	Brachytherapy
Average survival	0.50 [0.47–0.53]	0.75 [0.70–0.80]	0.6 [0.52–0.68]	0.6 [0.55–0.65]
Bowel problems	1	1	0.85	1
Incontinence	1	0.90	0.99	0.5
Erectile dysfunction	0.95	0.59	0.25	0.29

For example, the 95 % confidence interval for the average survival of patients under active surveillance is 9.4–10.6 years (Table 5.1), implying a confidence interval on the partial value from 0.47 to 0.53

Parameter uncertainty in weights is the variability in the estimation of a parameter of interest as a result of sampling. Although their underlying value may be the same, different decision makers will interpret a weighting scale differently and thus will come up with different weights. This can be reflected by calculating the mean weight along with the variance measure for each criterion over a group of decision makers. The parameter uncertainty is a function of the sample size and the underlying stochastic uncertainty. The larger the sample size n, the smaller the parameter uncertainty will be as it is a function of n with $1/\sqrt{n}$.

Individual weights are usually combined into an average weight over decision makers. The most commonly used method to combine individual weights is the arithmetic mean. However, in the analytic hierarchy process, the geometric mean is used to combine the weight estimates of different decision makers. The decision to use either the arithmetic mean or the geometric mean is important as it affects what method is appropriate for calculating the standard error around the mean weight.

Heterogeneity is the between-person variability that may be explained by the characteristics of the decision maker. For instance, erectile dysfunction as a result of treatment of prostate cancer may (or may not) be more important to a 40-year-old man compared to an 80-year-old man as the latter tends to have a less active sex life. It is important to have estimates of heterogeneity linked to background characteristics in MCDA, because the outcome of the analysis might be different for different (groups of) persons.

Stochastic uncertainty is the random, unexplained variability between different measurements of the weight estimates of one person. In most MCDA analyses, the magnitude of stochastic within-subject variability is not known as weight judgments are performed only once.

Heterogeneity is similar to stochastic uncertainty in that both cannot be reduced. The difference is that differences in weights as a result of heterogeneity of the subject need to be understood rather than minimized, while large random variability in weights is undesirable.

The choice of the weight elicitation technique induces structural uncertainty, as the use of different techniques can result in differences in weight estimates of the criteria, or may imply different (methodological) meanings of weights (Choo et al. 1999). Knowing that the exact weights vary based on the weight elicitation method stresses the need for sensitivity analysis on the final results. Previous research has shown that while exact weights might differ based on the weight elicitation method, the rank order of criteria is mostly maintained. In a few studies, it was shown that the differences in weights as a result of technique have a minor impact on the overall value of the alternatives. However, testing the range in which weights can vary before the rank order of alternatives changes (and to judge whether this extent of change is likely to happen as a result of the weight elicitation method) should be an important aim of sensitivity analysis (IJzerman et al. 2012a; van Til et al. 2014). To reduce structural uncertainty due to mismatches between the meaning of weights according to the MCDA model definitions and the decision makers' understanding of the weights' meanings, it is important to clearly explain the MCDA (elicitation) method to the decision makers.

Summarizing, all types of uncertainty influence the estimates of the weights. Parameter uncertainty can be made visible by presenting not only mean weights but also confidence intervals. Stochastic uncertainty and related structural uncertainty cannot be made explicit unless decision makers are asked to repeat their weight estimations with the same weight elicitation technique (stochastic uncertainty) or are asked to perform weight estimations with different weight elicitation techniques. Heterogeneity can be made visible by knowing and categorizing the decision makers and calculating mean weights (with confidence bounds) for the different subgroups.

5.4 Aggregation Methods

After the scoring and weighting steps are completed, performance values and criteria weights are (statistically) aggregated in an overall value. The most commonly used aggregation method is additive weighting, where the partial values on the different criteria are multiplied by their criteria weights and then summed up per alternative (see Chapter 4). The simplicity of additive weighting is attractive because it is easily understood by decision makers. From a theoretical perspective, other statistical aggregation methods might be preferred (see, e.g., Zhou and Ang 2009; Zanakis et al. 1998).

The choice of aggregation method is a form of structural uncertainty, since it can alter the model outcomes (Zhou and Ang 2009) and their interpretation. Moreover, because some approaches, such as the analytic hierarchy process, place very

specific requirements on the performance and weight elicitation techniques, the choice of aggregation method is a decision that has to be made early in the MCDA (Choo et al. 1999; Liberatore and Nydick 2008).

Another type of structural uncertainty is the decision at which point to aggregate the results of different decision makers in the weighting and performance stage. In essence, there are two ways to do so. One can average individual performance values and individual criteria weights (with measures of variance) and use an aggregation approach (for instance, an additive model) to calculate one overall value (with measures of variance). Alternatively, one can calculate an overall value for each individual and average the multiple estimations of overall value (with measures of variance). As aggregation is based on the product of two values, both approaches result in different average overall values and different measures of variance. Moreover, in the former case, providing a measure of parameter uncertainty by calculating a standard error of the overall value is difficult as the overall value is a sum of products of averages. One way to calculate the variance (and thus the standard deviation) of a product is the delta method (Rice 2006).

Finally, irrespective of the exact statistical aggregation method used, the output of an MCDA model is a point estimate of the overall value of the different alternatives. The impact of uncertainties on the aggregated overall value can be made explicit by calculating standard errors, confidence intervals, or ranges of the overall value of a treatment based on the standard errors (for parameter uncertainty) or standard deviations (for stochastic uncertainty) in the performances and weights. By reporting not only the point estimate of the overall value of a treatment but also its standard error or confidence interval, the parameter or stochastic uncertainty in the overall value is made visible.

5.5 Sensitivity Analysis

The outcome of a value-based MCDA method is an overall value for each alternative. However, without information on the uncertainty surrounding the weight estimates and performance values, the stability of the overall value is not known. Therefore, the confidence with which the results of the MCDA can be interpreted is then also not known. If one or multiple types of uncertainty are taken into account, this will result in a distribution of values around the point estimates. The shape and spread of the value distribution provide information about the stability of the conclusions that can be drawn from the analysis.

Sensitivity analysis is the study of the impact of uncertainty throughout the decision-making process on its outcomes. Structural uncertainty occurs as a result of the choices made in problem structuring with regard to the shape of the value tree, the type and number of criteria included in the analysis, and the MCDA method chosen to perform the analysis (including the weight elicitation and value performance method). The impact of structural uncertainty on the outcome can

only be made explicit by performing MCDA for the same problem with different value trees, criteria, and MCDA methods. This type of sensitivity analysis is a time-consuming process which is usually not performed.

A more common type of sensitivity analysis is studying the impact of parameter uncertainty in weights and performances or heterogeneity on the outcome(s) of an MCDA. When assessing the impact of uncertainty, one can do so throughout the whole MCDA process, identifying sources and measuring the amount of uncertainty at each stage separately and then studying its impact on the outcome(s) of the MCDA process. Alternatively, one can assess the impact of uncertainty on the overall value after the criteria weights and performance values are aggregated. Both are commonly termed "sensitivity analysis" in literature, and the latter is also sometimes termed "robustness analysis" or "post hoc sensitivity analysis." Although these two concepts are conceptually different, similar methods can be used during their application to demonstrate the uncertainty around the point estimates. In the next paragraphs, we will describe two commonly used methods for sensitivity analysis, namely, deterministic and probabilistic sensitivity analysis, and shortly touch upon some alternative methods.

In an earlier literature review in the healthcare context, 19 studies were identified where uncertainty was explicitly taken into account in the MCDA analysis (Broekhuizen et al. 2015a). In nine studies, the deterministic sensitivity approach was used, four studies used a probabilistic approach, and in other four studies (concerning environmental health issues), fuzzy set theory was applied. It seemed that in most MCDA-supported decisions, a deterministic sensitivity analysis was used because of its ease of use and because the increased insight in the stability of results was deemed sufficient. However, when the uncertainty in multiple model parameters needs to be considered simultaneously, approaches that use probability distributions should be applied.

Deterministic sensitivity analysis is the most straightforward method for (post hoc) sensitivity analysis. In deterministic sensitivity analysis, one parameter, that is a criterion weight or performance score, is varied at a time, and the impact of varying this parameter on the rank order of alternatives is observed. If the induced variation does not change the rank order of alternatives, i.e., the preference of one alternative over the other is preserved, the decision seems robust. Alternatively, one can assess the extent to which a parameter can be increased or decreased before the rank order of alternatives changes. The range in which the particular parameter is likely to change can be based on expert' judgments or the variation in available clinical data.

Recall that the urologists in the example took the confidence bounds for the average survival across treatments options from the literature. We already demonstrated in an earlier section that these can be transformed to confidence bounds on partial values. However, it might also be insightful to consider the impact of the range of partial values on the overall value of treatments. This can be done by inserting partial values for the lower and upper confidence bounds in the overall value function. This results in a confidence interval of the overall value in which the

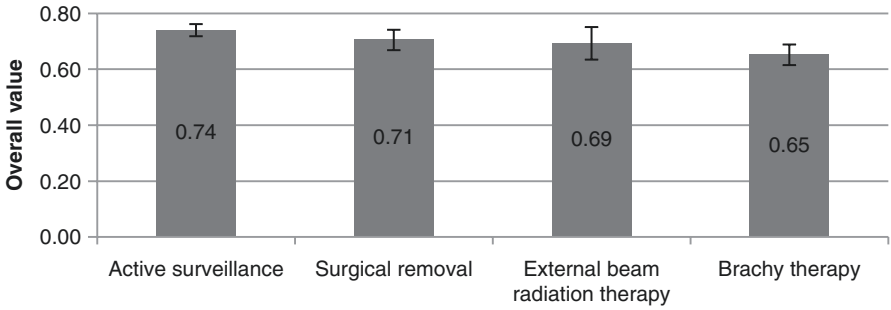

Fig. 5.3 Influence of parameter uncertainty in the survival estimates for the various treatments on their overall value in the urologists' example. Note the overlapping confidence bounds that indicate uncertainty regarding what treatment is more valuable

uncertainty depends only on the uncertainty present in the estimates for average survival (Fig. 5.3).

This deterministic sensitivity analysis reveals that there is overlap between the confidence bounds of overall values and that this depends (at least partly) on parameter uncertainty in the survival estimates. Furthermore, it seems that changing the survival estimates within confidence bounds can lead to rank reversals of alternatives. For example, it is possible that external beam radiation therapy has a higher overall value than surgical removal. The question remains, however, how likely it is that such a rank reversal between external beam radiation therapy and surgical removal occurs.

Deterministic sensitivity analysis can also be used to assess the impact of (uncertainty in the) criterion weights on the alternatives' overall value by manually varying the criterion weights one by one and observing how the overall values of the alternatives change. For example, if we increase the weight of survival and thereby decrease the weights of the other criteria (because weights add up to one), alternatives that have longer survival will increase in overall value compared to alternatives with a relatively shorter survival. One can vary each criterion weight from its lowest possible value to its highest possible value and observe the effect on overall value of the alternatives (Fig. 5.4). Alternatively, and more effectively, one can vary the weights within the confidence bounds resulting from parameter uncertainty and heterogeneity in preferences within the group and see whether this variation would change the outcome of the model.

Another particular deterministic sensitivity graph, popular in health-economic assessments, is the tornado graph. A tornado graph shows the impact on model outcomes of arbitrary fixed changes (e.g., −10 % and +10 %) in single model parameters. It is especially useful for determining which model parameter has the greatest influence on the outcome (Briggs et al. 2012).

The results of the analysis as presented in Fig. 5.4 will provide the urologist with more information on the robustness of their results. Active surveillance has the highest value when the mean criteria weights of the group of urologists are

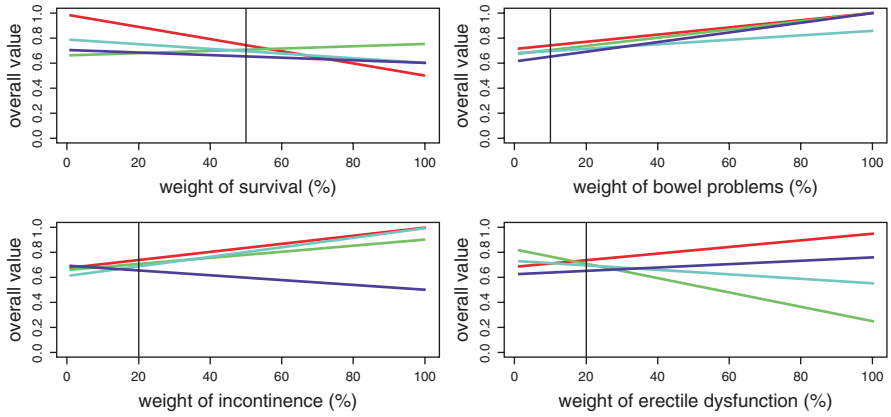

Fig. 5.4 Each of these three graphs shows how the overall value of treatments would vary, had the urologists chosen different (i.e., higher or lower) weights for each of the criteria. This overall value is on the vertical axes, the weights are on the horizontal axes, and the vertical black lines denote the point estimates of the weights. *Red* active surveillance, *Green* surgical removal, *Light blue* external beam radiation therapy, *Purple* brachytherapy

used. However, intersections between lines imply rank reversals between treatments based on changes in criteria weights. If, for example, the weight put on survival would increase above 58 %, surgical removal would have the highest value, making it the preferred treatment. However, the threshold, i.e., the weight for survival where a rank reversal between surgical removal and active surveillance occurs, is 8 % removed from the initial point estimate, and this falls within the variation of weights given by the individual urologists but not within the confidence interval of the average weight. The urologists must determine whether such an increase in weight is likely. For now, based on the deterministic sensitivity analysis, it seems that rank reversals are unlikely and that the preference for active surveillance is robust.

Although it is easy to implement, deterministic sensitivity analysis has two important drawbacks. First of all, only one model parameter (weight or performance score) is varied at a time. This is unrealistic because it assumes uncertainty in only one parameter, while actually multiple (or all) model parameters can be uncertain. Second, manually changing of model parameters, such as presented above or in a tornado graph, does not take into account the actual uncertainty in the model parameters. For example, if the observed range of a weight is between 40 % and 60 %, it does not make sense to investigate rank reversals that occur when the weight is 80 %. This is a relevant issue for the urologists, because their deterministic sensitivity analysis shows them that rank reversals occur at particular combinations of survival estimates, but they cannot quantify the likelihood with which this might happen.

Instead of a deterministic sensitivity analysis, a probabilistic approach could be used to gain insight into the impact of the combination in uncertainty in the clinical evidence, scores, and/or weights on the overall value of the alternatives. For

example, in a study by Wen et al., two different methods (a delta-method approach and a Monte Carlo approach) for constructing a confidence interval of the overall benefit-risk score from an MCDA model were compared (Wen et al. 2014). The objective of the study was to provide suggestions for incorporating the uncertainty in performance data based on clinical evidence into the MCDA model when evaluating the overall benefit-risk profiles of different treatment options. In a study by Broekhuizen et al., the impact of uncertainty in the performance estimates based on clinical evidence was studied along with the uncertainty in criteria weights as given by patients (Broekhuizen et al. 2015b). In stochastic multi-criteria acceptability analysis (SMAA), uncertainties in preference data and clinical trial data are combined, and a non-informative (uniform) distribution based on the rank order of criteria is used for the weight distributions (Tervonen et al. 2011; van Valkenhoef et al. 2012). Finally, Caster et al. use qualitative data on the rank order of criteria and combine this with probability distributions for clinical data (Caster et al. 2012).

In a probabilistic approach, uncertainty in model parameters is represented with probability distributions. There are many different types of probability distributions. When data is available, the empirical distribution can be used or assumptions with regard to a parametric distribution must be made. A comprehensive review of methods for eliciting probability distributions from (groups of) experts can be found in (O'Hagan et al. 2006).

After selecting or eliciting a probability distribution that reflects the uncertainty in each model parameter, one can assess how the uncertainty in all these parameters translates to uncertainty in the overall value of the alternatives, for example, by means of a Monte Carlo simulation approach. This approach consists of sampling from the distributions of one or multiple model parameters simultaneously and then calculating the overall value of the alternatives for each of these (combinations of) sampled estimates. By repeating this process a large number of times (e.g., 1000 or more), it can give decision makers an idea about the likely distribution of overall value of each included alternative (Broekhuizen et al. 2015b).

In our prostate cancer example, a normal probability distribution is selected for survival because of the large sample sizes in the clinical trials. After parametrizing this distribution based on standard errors reported in the clinical paper and running the Monte Carlo simulations, the distributions for the overall values as presented in Fig. 5.5 are obtained.

The amount of overlap between these distributions is an indicator of the likelihood that the treatments are in the correct preference order, while the width of the curves is an indicator of how likely the point estimates of the values are. If there is much overlap between the value distributions of two treatments, and the value distributions are "wide" (such as with the light-blue line), there is more uncertainty about which treatment has the highest value. This uncertainty can be quantified by taking the percentage of Monte Carlo samples in which a particular treatment has the highest value. This is called the first ranking probability. One minus the first ranking probability is a surrogate of decision uncertainty as it estimates the probability that the alternative with the highest mean value does not have the highest rank (Table 5.3).

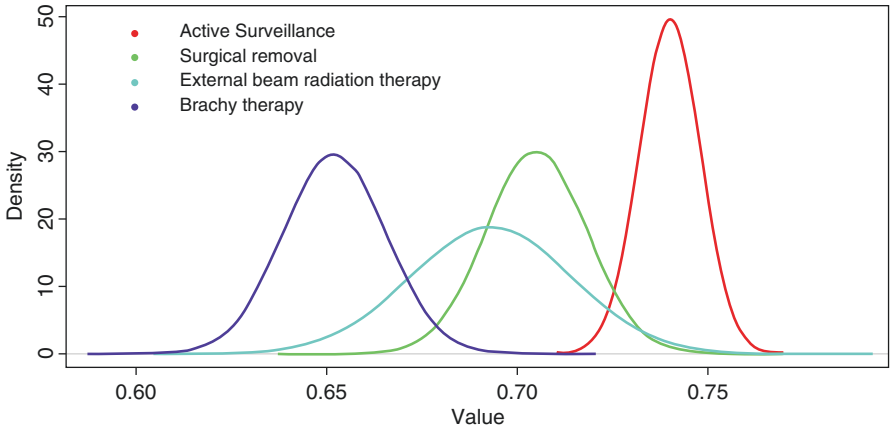

Fig. 5.5 Distribution of overall value of the alternatives in the urologists' case, when normal distributions based on clinical literature are assigned to the "survival" performance parameter. Estimated using a gaussian kernel density

Table 5.3 Ranking probabilities in the urologists' case, when the uncertainty in the survival estimates is represented with normal distributions (see Fig. 5.5) and after running 5000 Monte Carlo simulations

Probability of…	Active surveillance	Surgical removal	External beam radiation therapy	Brachytherapy
… being ranked 1st	97 %	1 %	2 %	–
… being ranked 2nd	3 %	68 %	29 %	–
… being ranked 3rd	–	31 %	64 %	5 %
… being ranked 4th	–	–	5 %	95 %

Please note that only uncertainty in survival is taken into account. When more parameters would be assigned probability distributions to reflect uncertainty, the probability of rank reversals may increase

Other approaches apart from deterministic and probabilistic approaches to incorporate uncertainty analysis in MCDA have been identified, namely, Bayesian frameworks, fuzzy set theory, and grey theory (Broekhuizen et al. 2015a).

Within the Bayesian framework, a distinction can be made between approaches based on Bayesian networks (Fenton and Neil 2001) and approaches based on Dempster-Shafer theory (Beynon et al. 2000). Fuzzy set theory aims to capture the ambiguity present in human language and judgment and is often combined with the AHP method of MCDA. Comparable to fuzzy set theory are approaches based on grey theory (Ju-Long 1982). With these approaches one can address all types of uncertainty except for structural uncertainty. The applicability of these methods for addressing uncertainty is sometimes strictly dependent on the specific form of MCDA used. For example, SMAA is a strictly probabilistic method (Lahdelma and Hokkanen 1998). Other MCDA methods like AHP, PROMETHEE, TOPSIS, and ELECTRE can be combined with (almost) all uncertainty approaches.

5.6 Summary and Conclusions

Uncertainty is introduced into an MCDA at the following stages: the problem structuring stage, the performance valuation stage, and the criteria weighting stage. Structural uncertainty is introduced as a result of methodological choices such as the MCDA method, structuring of the value tree, type of weight and performance elicitation techniques used, and aggregation method. Parameter uncertainty occurs because of sampling error. Stochastic uncertainty is the uncertainty as a result of random, unexplained variation and can be made visible by presenting, e.g., histograms/densities for the weights and performance values. Heterogeneity is the explained variation as a result of different background characteristics and values of the respondents. Often it is not possible or even desirable (in the case of heterogeneity) to reduce uncertainty, but the aim of this chapter was to explain how uncertainty can be made explicit throughout the decision process and to study its influence on the output of the MCDA.

We emphasized how the quantitative outcomes of the model depend on the way in which weight and performances are aggregated, while the interpretation of the output of the model depends also on the way in which the output is presented to the decision maker.

Through sensitivity analysis, the impact of uncertainty on the outcome of the decision analysis can be made explicit. We have demonstrated how different types of uncertainty in the inputs of the MCDA model influence its outputs and how uncertainty can be quantified with different measures of variability (standard deviation, standard error, range) or can be graphically displayed. Both deterministic sensitivity analysis and probabilistic sensitivity analysis were explained.

To fully analyze the impact of uncertainty in MCDA, additional efforts may be required from the decision makers in terms of additional model inputs (measures of variation, probability distributions, ranges of weights or scores) and from decision analyst in terms of analytic skills. A balance must be struck between increasing confidence of the decision makers in the output of the MCDA by demonstrating the impact of uncertainty and not losing confidence of decision makers in the MCDA itself by making the analyses too complicated.

References

Belton V, Pictet J (1997) A framework for group decision using a MCDA model: sharing, aggregating or comparing individual information? J Decis Sys 6(3):283–303. Available at: http://www.tandfonline.com/doi/abs/10.1080/12460125.1997.10511726. Accessed 18 Feb 2014

Belton V, Stewart TJ (2002) Multiple criteria decision analysis: an integrated approach, 2nd edn. Kluwer Academic Publishers, Dordrecht

Beynon M, Curry B, Morgan P (2000) The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling. Omega 28(1):37–50

Bojke L, Soares M (2014) Decision analysis: eliciting experts' beliefs to characterize uncertainties. Elsevier Ltd. Available at: http://www.sciencedirect.com/science/article/pii/B9780123756787014061
Bojke L et al (2010) Eliciting distributions to populate decision analytic models. Value Health 13(5):557–564
Bouyssou D (1990) Building criteria: a prerequisite for MCDA. In: Readings in multiple criteria decision aid. Springer, Berlin, pp 58–80
Briggs AH et al (2012) Model parameter estimation and uncertainty: a report of the ISPOR-SMDM modeling good research practices task force-6. Value Health 15(6):835–842. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22999133
Broekhuizen H et al (2015a) A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Pharmacoeconomics 33(5):445–455
Broekhuizen H et al (2015b) Estimating the value of medical treatments to patients using probabilistic multi criteria decision analysis. BMC Med Info Decis Making 15(1):1–10. Available at: http://dx.doi.org/10.1186/s12911-015-0225-8
Caster O et al (2012) Quantitative benefit-risk assessment using only qualitative information on utilities. Med Decis Making 32(6):E1–E15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22936214 [Accessed 3 Dec 2012]
Choo EU, Schoner B, Wedley WC (1999) Interpretation of criteria weights in multicriteria decision making. Comput Ind Eng 37(3):527–541. Available at: http://linkinghub.elsevier.com/retrieve/pii/S036083520000019X
Cooperberg MR et al (2012) Primary treatments for clinically localised prostate cancer: a comprehensive lifetime cost-utility analysis. BJU Int 111(3):437–450
Bana E Costa, C. A, De Corte J.-M, Vansnick J.-C (2012) Macbeth. Inter J Info Tech Decision Making 11(02):359–387. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219622012400068
Durbach IN, Stewart TJ (2012) Modeling uncertainty in multi-criteria decision analysis. Eur J Oper Res 223(1):1–14. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221712003591. Accessed 27 Jan 2014
Fenton N, Neil M (2001) Making decisions: using Bayesian nets and MCDA. Knowl Based Syst 14(7):307–325. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0035505967&partnerID=tZOtx3y1. Accessed 18 Feb 2014
French S (1995) Uncertainty and imprecision: modelling and analysis. J Oper Res Soc 46(1):70–79. Available at: http://www.jstor.org/stable/2583837
Hayes J et al (2013) Observation versus initial treatment for men with localized, low-risk prostate cancer. Ann Intern Med 158(12):853–860
IJzerman MJ, Van Til JA et al (2012) A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. Patient 5(1):45–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22185216
IJzerman MJ, Til JA, Snoek GJ (2012) Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. Patient Patient-Centered Out Res 1(4):265–272. Available at: http://dx.doi.org/10.2165/1312067-200801040-00008
Ju-Long D (1982) Control problems of grey systems. Syst Control Letters 1(5):288–294. Available at: http://linkinghub.elsevier.com/retrieve/pii/S016769118280025X
Lahdelma R, Hokkanen J (1998) SMAA-stochastic multiobjective acceptability analysis. Eur J Oper 2217(97):137–143. Available at: http://www.sciencedirect.com/science/article/pii/S037722179700163X. Accessed 26 Mar 2012
Liberatore MJ, Nydick RL (2008) The analytic hierarchy process in medical and health care decision making: a literature review. Eur J Oper Res 189(1):194–207
O'Hagan A et al (2006) Uncertain judgements: eliciting expert probabilities. John Wiley & Sons Ltd, Chichester. Available at: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470029994.html
Phillips LD (1999) Group elicitation of probability distributions: are many heads better than one? In: Shanteau J, Mellors B, Schum D (eds) Decision science and technol- ogy: reflections on the

contributions of ward edwards. Kluwer Academic Publishers, Norwell, p 313–330. Available at: http://eprints.lse.ac.uk/10520/

Rice J (2006) Mathematical statistics and data analysis. Thomson Brooks/Cole Belmont, CA, USA

Taner M, Sezen B, Antony J (2007) An overview of six sigma applications in healthcare industry. Int J Health Care Qual Assur 20(4):329–340

Tervonen T et al (2011) A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med 30(12):1419–1428. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21268053. Accessed 26 Mar 2012

van Til J et al (2014) Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework. Cost Eff Res Allocation 12:22

van Valkenhoef G et al (2012) Multicriteria benefit-risk assessment using network meta-analysis. J Clin Epidemiol 65(4):.394–403. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22197518

Wen S, Zhang L, Yang B (2014) Two approaches to incorporate clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. Value Health 17(5):619–628. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25128056. Accessed 21 Aug 2014

Zanakis SH et al (1998) Multi-attribute decision making: a simulation comparison of select methods. Eur J Oper Res 107(3):507–529

Zhou P, Ang BW (2009) Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. Social Indic Res 94(1):83–96