Methodological aspects of multi-criteria decision analysis for policy support:

A case study on pharmaceutical removal from hospital wastewater

Schuwirth, N.[1]*, Reichert, P.[1], Lienert, J.[1]

[1]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, CH-8600 Dübendorf, Switzerland

*Corresponding author: Tel.: +41 58 765 5528, Fax: +41 58 765 53 75, E-mail: nele.schuwirth@eawag.ch (N. Schuwirth)

Abstract

Decision making in public and political contexts can be complex. Multi-attribute value/utility theory (MAVT/MAUT) can support such decision processes by providing a transparent framework that helps focusing on objectives and corresponding degrees of achievement by different alternatives.

Eliciting preferences with MAVT/MAUT can be time consuming and cognitively challenging. Therefore, it might not be feasible to elicit full preference functions with standard methods. To deal with this problem, we suggest a simplified elicitation procedure that combines (a) the elicitation of values instead of utilities at lower-levels of the objectives hierarchy and conversion to utilities to consider risk attitudes at appropriate higher levels, (b) the use of linear value functions for sub-objectives with minor effects on the overall value, and (c) sensitivity analyses to check the robustness of results regarding these assumptions and the elicitation process in general. Furthermore, we developed a modified Swing procedure ("Reversed Swing") to elicit weights for cases in which the hypothetical alternatives of the conventional Swing technique are unrealistic.

We applied this procedure to a case study on pharmaceutical removal from wastewater of a typical Swiss hospital involving 13 stakeholders. Aim of the interdisciplinary research group was to assess a large bundle of combinations of novel point source measures. The ultimate policy objective was to develop consensus solutions which are acceptable to all important stakeholder groups. We hope that the suggested simplified procedure stimulates the application of transparent and conceptually satisfying decision support

32 methods in environmental management, which is needed to justify policy decisions to the
33 public.

1. Introduction

35 Aims of multi-criteria decision analysis (MCDA) for policy support are to facilitate a fair
36 discussion about different management options, to address different points of view of
37 stakeholders, and to provide information about consequences of different options. This
38 includes the quantification of the prediction uncertainty and the valuation of
39 consequences by stakeholders. Furthermore, by allowing for an analysis of the causes
40 of poor valuation results of certain alternatives for certain stakeholder groups, MCDA
41 should help to stimulate the process of creating new options that are acceptable to all.
42 The ultimate goal may be to guide an equitable process for finding an alternative that - to
43 the best scientific knowledge and to an acceptable degree - fulfills the objectives of
44 various stakeholders and for transparently communicating the reasons for this decision.
45 Policy decisions are especially difficult if they are embedded in complex institutional
46 settings and involve many parties with different perspectives. Further challenges are
47 high uncertainties in the prediction of consequences and conflicting objectives. MCDA
48 techniques are very suitable to deal with these problems in a structured way. This is of
49 particular importance in cases where stakes are high and/or the decision is intellectually
50 difficult, or the value elicitation results or the consequences of alternatives are uncertain
51 (Von Winterfeldt and Edwards, 1986). Systematic reviews of application fields of
52 decision analysis between 1970 and 2001 reported in operations research journals are
53 given in Corner and Kirkwood, (1991) and Keefer et al., (2004). Kiker et al., (2005)
54 review MCDA applications to environmental decision making.
55 In this article, we report about a test of the applicability of MCDA in a real case study
56 that includes the challenges discussed above. The decision concerns the introduction of
57 point-source measures that decrease pharmaceutical concentrations in the wastewater
58 of a typical Swiss hospital. Today, pharmaceuticals are not fully eliminated in wastewater
59 treatment plants; their presence in water bodies is giving rise to concern (e.g., Ankley et
60 al., 2007; Kolpin et al., 2002). Besides upgrading treatment plants (Joss et al., 2008), a
61 possible approach is to reduce the pharmaceutical load coming from important point
62 sources, such as hospitals. While technical-engineering options are being intensively
63 researched (e.g., Beier et al., 2010; Lenz et al., 2007, www.pills-project.eu), including
64 stakeholders in the decision process has so far mostly been neglected. However,

65 societal acceptance is crucial to policy makers who must decide whether (public tax)

66 investments to remove pharmaceuticals are worthwhile. In our interdisciplinary project

67 we involved 13 stakeholders and evaluated 68 combinations of technical alternatives

68 based on nine different objectives. The more technical results are presented in Escher et

69 al., (2011) and Lienert et al., (2011). In this article, we focus on the methodological

70 challenges of the MCDA procedure.

71   To support this decision, we chose multi-attribute value and utility theory (MAVT/MAUT)

72 due to the following four properties: Since it is founded on fundamental axioms of

73 rational choice (Savage, 1954; Von Neumann and Morgenstern, 1947; Von Winterfeldt

74 and Edwards, 1986), results are justifiable, which is important for policy decisions that

75 have to be defended in the public arena. Second, the uncertainty of predictions and the

76 risk attitudes of the decision makers or stakeholders can be handled with MAUT. This is

77 again very important for complex policy decisions, where the consequences of different

78 alternatives (e.g. management options) can not be predicted precisely. Third, it can deal

79 with a large number of alternatives without an increase of the elicitation effort compared

80 to a study with a smaller number of alternatives. The reason is that value and/or utility

81 functions are elicited from the decision maker or stakeholder independently from the

82 alternatives based on his or her preferences about the fulfillment of the different

83 objectives. All that needs to be known for preference elicitation is the range of each

84 attribute over all alternatives, i.e., the best- and worst-possible level of each attribute.

85 The elicited value and/or utility function are then used to evaluate the alternatives.

86 Therefore, the number of alternatives is irrelevant for the elicitation procedure and any

87 number of additional alternatives can be introduced at later stages as long as their most-

88 extreme outcomes stay within the ranges of the attributes defined for preference

89 elicitation. Hence, the fourth argument for using MAVT/MAUT is that we want to be able

90 to include new alternatives at any stage of the decision procedure (Keeney, 1992)

91 without the need for re-eliciting preferences and without changing the ranking of the old

92 alternatives. The creation of such new alternatives, in particular alternatives with a higher

93 consensus potential, can be stimulated by the decision analysis and might even be one

94 important objective of the whole process.

95   Areas where MAVT and/or MAUT have been applied to environmental decision making

96 include forest management (reviewed by Ananda and Herath, 2009, Hayashida et al.

97 2010), natural resource management (reviewed by Mendoza and Martins, 2006),

98 different fields of water management (reviewed by Joubert et al., 2003; Linkov et al.,

99 2006) and river management (e.g., Corsair et al., 2009; Reichert et al., 2007; Ríos-Insua

100 et al., 2006).

101 Other commonly used decision support methods do not exhibit all of the four properties

102 discussed above and were therefore not considered in this study. This applies to

103 outranking procedures (PROMETHEE (Brans et al., 1986; Brans and Vincke, 1985), see

104 discussion by Mareschal et al., 2008), ELECTRE (Roy, 1996), see discussion by Wang

105 and Triantaphyllou, 2008), and the Analytic Hierarchy Process (AHP (Saaty, 1980), see

106 discussion of Bana E Costa and Vansnick, 2008; Dyer, 1990; Perez et al., 2006).

107 Many textbooks introduce MAVT and/or MAUT, (e.g. Belton and Stewart, 2002;

108 Clemen, 1996; Eisenführ et al., 2010; Keeney and Raiffa, 1993). However, eliciting

109 preference functions for complex decisions with many objectives is intellectually

110 challenging and time consuming. Important policy stakeholders usually have very limited

111 time, and it may be necessary to simplify the elicitation task (e.g. Janssen, 2001; Stewart,

112 1995). However, elicitation is prone to cognitive biases (Borcherding et al., 1991;

113 Hamalainen and Alaja, 2008; Krzysztofowicz and Duckstein, 1980; Weber and

114 Borcherding, 1993; Weber et al., 1988). This problem is aggravated by the introduction

115 of inadequate simplifying assumptions that could further distort the modeled stakeholder

116 preferences and produce deceptive results.

117 The intent of this paper is to simplify MAVT/MAUT-methodologies in such a way that

118 they become better applicable in multi-faceted decision problems in a public context,

119 where the textbook procedures reach their limits. We strongly believe that further

120 development in this direction is needed rather than to compromise on the validity of

121 procedures, which may result in serious violations of the axioms of decision theory. We

122 therefore worked on a conceptually satisfying elicitation procedure that is doable in face

123 to face interviews within a time frame of a few hours. We demonstrate its applicability in

124 the case study. It deals with high prediction uncertainty and shows how the influence of

125 simplifications can be evaluated by sensitivity analyses.

126 In section 2 we describe methodological aspects of the decision support approach. We

127 present the case study and major results in sections 3 and 4. In section 5 we discuss the

128 applicability of the elicitation procedure and draw conclusions in section 6.

2. Decision Support Methodology

130 A decision support process for policy advice based on MAVT/MAUT can be structured

131 according to Figure 1 (modified after e.g. Clemen, 1996; Eisenführ et al., 2010). For

clarity, we only drew the main arrows, more iterations may be appropriate. A particular challenge is the elicitation of the preference function (step F in Fig 1). We therefore focus on performing this task with minimizing the time requirements for individual stakeholder elicitations. In particular, we assume that objectives have already been elicited and structured hierarchically (step C in Fig. 1) and that the quantitative preference elicitation should be based on this hierarchy.
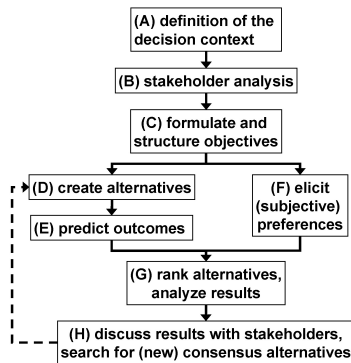

Fig. 1: Scheme of the decision support process

2.1 Challenges of preference elicitation

Empirical research indicates systematic deviations of individuals from rational behavior as formulated in MAVT/MAUT in actual intuitive decision making (e.g. Allais, 1953, summarized in Eisenführ et al., 2010). This led to the development of descriptive preference theories that account for this behavior (e.g. Birnbaum, 2008; Gregory and Keeney, 2002; Kahneman and Tversky, 1979). Despite these deviations of actual decision making from rational behavior as formulated by MAVT/MAUT, rational decision theory may still be a good choice as a basis for public decision support as its purpose in this application area is to support decision making that can be justified with rational arguments. Despite the use of rational decision theory, the elicitation procedure must be designed carefully to avoid biases investigated in behavioral decision research. Below, we suggest ways to handle critical steps.

When considering uncertainty in the prediction of outcomes, the risk attitude of the decision maker has to be included to rank alternatives. This requires eliciting a utility function. To disentangle strength of preference and (relative) risk attitudes (Dyer and Sarin, 1982), it is often recommendable to elicit value functions and switch to utilities at an adequate hierarchy level of the objectives hierarchy. This has the additional advantage that it makes it possible to elicit the preferences regarding certain branches of

158 the objective hierarchy from experts (who may be better in assessing the degree of

159 fulfillment of sub-objectives that may be based on technical attributes) without having to

160 adopt their risk attitudes. The risk attitudes and trade-offs to other branches at higher

161 levels of the objectives hierarchy can then be elicited from stakeholders or decision

162 makers (Abbas, 2010; Barron et al., 1984; Dyer and Sarin, 1979; Keeney and Raiffa,

163 1993; Sarin, 1982).

164   The elicitation procedure can be decomposed as follows:

165   (I)   Elicit value or utility functions for lowest level sub-objectives.

166   (II)   Determine appropriate aggregation methods for values or utilities at higher

167   hierarchical levels; elicit aggregation parameters.

168   (III) If values were elicited, convert them to utilities at appropriate levels of the

169   objectives hierarchy.

170 This procedure has the additional advantage that eliciting value functions is much easier

171 and less prone to cognitive biases since one has only to compare different outcomes

172 and not lotteries of outcomes. Furthermore the predicted probability distributions of

173 values offer additional insight to expected utilities.

174 **2.2 Elicit value or utility functions for lowest level sub-objectives**

175 Value functions measure the degree of fulfillment of an objective on a scale between 0

176 and 1. Measurable (i.e., interval scale) value functions are required for quantifying trade-

177 offs in the next step. Single-attribute value functions can be elicited with standard

178 methods (e.g., Fishburn, 1967; Von Winterfeldt and Edwards, 1986). We find the

179 Midvalue Splitting (Bisection) method most appropriate. It leads to more reliable results

180 than Direct Rating and might be easier to apply than the Difference Standard technique.

181 Crucial parts of elicitation are consistency checks and some training to avoid the goal-

182 directed bias (e.g., Martin et al., 2000). Otherwise, interviewees often state that the

183 midvalue point is close to the preferred endpoint because they do not focus on

184 improvements but on final outcomes.

185   Many applications assume that the degree of fulfillment of the objective depends

186 linearly on the attribute level. This assumption is certainly not valid for the many

187 examples with decreasing marginal values (e.g., for income). However, in decisions with

188 many objectives, proper elicitation of value functions for all sub-objectives might be

189 unrealistic. As a compromise, we suggest to elicit value functions only for the most

190 important sub-objectives, assuming linear functions for those which have a significantly

191 smaller influence on the overall value. The influence of this simplification must be

192 checked by sensitivity analyses a posteriori.

193   The most common methods to directly elicit single-objective utility functions are the

194 Certainty Equivalent and the Variable Probability methods (Eisenführ et al., 2010; Von

195 Winterfeldt and Edwards, 1986). However, both techniques compare lotteries with

196 certain outcomes. We see this as disadvantage, since people facing this extreme choice

197 often intuitively violate the independence axiom (Allais Paradox; Allais, 1953). In policy

198 decisions, usually all consequences of alternatives have some uncertainty. We therefore

199 suggest using elicitation procedures where two lotteries are compared such as the

200 Trade-Off technique (Bleichrodt et al., 2001) or simplifications of this technique.

2.3 Aggregation of value/utility functions

201

202   The by far most popular aggregation technique is additive aggregation, i.e., that the

203 value v at a higher hierarchical level is the weighted average (or weighted arithmetic

204 mean) of the values v_i of the sub-objectives.

205
$$v_{\text{add}} = \sum_{i=1}^{n} w_i v_i \qquad (1)$$

206   Here, v_{add} is the aggregated value, v_i is the result of the value function characterizing

207 the fulfillment of the sub-objective i at the lower level dependent on attributes, w_i is the

208 weight of the sub-objective i. The weights sum to unity.

209   Additive aggregation implies that a low value of one sub-objective can be compensated

210 by large values of other sub-objectives. Therefore, this aggregation technique must fulfill

211 relatively strong independence conditions (Dyer and Sarin, 1979; Keeney and Raiffa,

212 1993), which must be verified in each case. It is useful to decompose objectives into

213 sub-objectives such that objectives are preferential-independent and the assumptions of

214 additive aggregation are approximately fulfilled, whenever this is possible.

215   In cases where the compensation effect of additive value functions is not appropriate,

216 other aggregation methods must be used. Examples are multiplicative aggregation (Dyer

217 and Sarin, 1979; Keeney and Raiffa, 1993), requiring weaker independence conditions,

218 the Cobb-Douglas function (Cobb and Douglas, 1928; Varian, 2010) commonly used in

219 microeconomics, which is the weighted geometric mean, worst case aggregation, or

220 mixtures. We found few applications of these in complex decision support studies (but

221 see Duckstein et al., 1994; Keeney and Wood, 1977; Raju and Vasan, 2007; Torrance et

222 al., 1996) although we think that there are cases in which approximate additivity cannot

223   be achieved through a good choice of sub-objectives (e.g. if the sub-objectives describe

224   complementary properties that should all be achieved to get a good fulfillment of the

225   overarching objective). In addition to finding an adequate aggregation function,

226   parameter elicitation of such more complex aggregation methods needs more research.

2.4 Elicitation of weights, the Trade-Off and Swing methods

228   Two standard techniques to estimate the parameters of additive aggregation, the

229   weights w_i, are the Trade-Off and the Swing methods, described e.g. in Eisenführ et al.,

230   (2010). In the Trade-Off method, weights are calculated by searching indifference

231   between outcomes that differ in two attributes. The literature and own experience

232   indicates that the questions are rather difficult to answer (Borcherding et al., 1991), but

233   comparatively reliable and theoretically most defensible (Martin et al., 2000). If there are

234   many objectives, this method might be too time-consuming. The Swing method uses a

235   reference state in which all attributes are at their worst level and let the interviewee

236   assign points to states in which one attribute moves to the best state. The weights are

237   then proportional to these points. This method is often used (e.g. Von Winterfeldt and

238   Edwards, 1986). It is fairly fast and interviewees readily give answers. This is an

239   advantage, but holds the risk that people respond without thoroughly considering the

240   consequences of their answers. Another advantage of the Swing method is that it does

241   not depend on the shape of the value functions of the sub-objectives. Only the attribute

242   ranges must be known and the levels of the best and worst outcomes (in most cases

243   corresponding to the endpoints of the ranges). This makes it possible to elicit weights

244   prior to assessment of the value functions of the sub-objectives, which can reduce the

245   splitting bias, as mentioned below. The disadvantages are that the technique is based

246   on direct rating, it does not include consistency checks, and the extreme outcomes to be

247   compared may not correspond to a realistic alternative, which makes the questions

248   difficult to answer. However, in principle it is also possible to use any two reference

249   points, and not only the extremes (Belton and Stewart, 2002).

2.5 Elicitation of weights, the Reversed Swing method

251   To extend the Swing technique to situations in which the state with all attributes at their

252   worst level is not realistic, we developed the "Reversed Swing method". Contrary to the

253   classical method, the reference state has all attributes at their best level and the

254   outcomes to evaluate have the attribute(s) of one sub-objective at the worst level. The

255 best outcome (all attributes at best level) receives 100 points; the outcome with the most

256 important attribute at the lowest level gets zero points. Note: this does not correspond to

257 a value of zero, since all but one of the attributes are at the best level. This has to be

258 discussed explicitly with the interviewees to avoid a biased evaluation. The interviewee

259 has then to specify scores, t_r, between 0 and 100 to the other outcomes with the

260 attribute(s) of one sub-objective r at the worst level.

261 The weights, w_r, of the additive aggregation scheme given by Eq. 1 can easily be

262 calculated from the scores, t_i, of the Reversed Swing method:

263
$$w_r = \frac{100 - t_r}{\sum_{i=1}^{m}(100 - t_i)}$$
(2)

264 The procedure is illustrated by the tool we developed for the interviews of the case

265 study described below. It is given in the appendix (p. 2). We also show the tool for the

266 classical Swing method (appendix p. 3). To illustrate the differences and similarities

267 between the two approaches, we walk through an example (appendix p. 4). We

268 recommend Reversed Swing if the hypothetical outcome combinations are more realistic

269 than those of the classical Swing method, as in our case study. Torrance et al., (1996)

270 used a similar approach, but we found the introduction of disutilities (one minus utility) an

271 unnecessary complication. Explicit and repeated discussions of the attribute ranges

272 during the Swing (or Reversed Swing) procedure help interviewees to adjust weights to

273 the attribute range. This avoids distortions due to the range effect (Von Nitzsch and

274 Weber, 1993).

275 Empirical studies show that objectives receive higher weights if they are split into more

276 detailed levels (splitting or overweighting bias; (Borcherding et al., 1991; Hamalainen

277 and Alaja, 2008; Weber and Borcherding, 1993; Weber et al., 1988). To avoid this

278 common bias, we suggest to hierarchically aggregate the objectives and to elicit the

279 ranking of the main objectives before presenting the sub-objectives. However, this

280 constrains the elicitation methods to those that do not require value functions for the

281 lowest level sub-objectives in advance. If monotonic value functions can be assumed,

282 the classical or Reversed Swing method are appropriate since they deal with swings

283 over the whole attribute range, and the values for the worst and best attribute level are

284 zero and one. For the reasons mentioned above, we strongly recommend to use the

285 Trade-Off method, at least as consistency check for the weights of the most important

286 attributes (appendix p. 5 shows an elicitation tool). This consistency check can be

287 performed after having elicited both, weights and value functions at the lowest sub-
288 objective level.

2.6 Convert values to utilities at appropriate levels of the hierarchy

290 If values were elicited for some branches of the hierarchy and the predictions are
291 uncertain, they must be converted to utilities (Abbas, 2010). As utilities can be applied to
292 certain outcomes also, compatibility requires that values and utilities share the same iso-
293 surfaces as a function of the attributes (in the case of n attributes, these are n-1
294 dimensional manifolds in the attribute space). This implies that the utility can be
295 expressed as a function of value only and does not have to be elicited starting again
296 from the attributes. States of given value to be compared when eliciting utilities should,
297 however, be communicated by underlying attributes to avoid problems similar to those of
298 direct rating procedures of value elicitation. There is some freedom in choosing these
299 attributes as the utility cannot depend on multiple representations of the same value by
300 different combinations of attributes.
301 Instead of eliciting risk attitudes explicitly, an a posteriori sensitivity analysis to simple
302 parameterizations of risk attitudes can be performed (e.g. exponential dependence of
303 utility on value, Keeney and Raiffa, 1993, p.330). If the resulting ranking of alternatives
304 does not depend on the risk attitude (when it is varied within a reasonable interval),
305 elicitation of risk attitudes is not necessary for the particular decision problem.

3. Case study design

3.1 Outline of the case study

308 Next, we illustrate such a decision process with an exemplary case study. It was
309 conducted for the typical cantonal hospital in Baden (canton Aargau) in the German
310 speaking part of Switzerland to support the decision about how to deal with the hospital
311 wastewater (the federal state of Switzerland consists of 26 territorial member states
312 termed cantons).
313 As mentioned above, hospitals are potentially important point dischargers of
314 pharmaceuticals. In our case, the wastewater from the cantonal hospital Baden
315 contributes by ca. 38% to the total load of pharmaceuticals at the communal wastewater
316 treatment plant (Escher et al., 2011; Lienert et al., 2011). Various European policy
317 makers and researchers (e.g., www.pills-project.eu) are currently debating whether on-
318 site treatment of hospital wastewater is feasible to decrease the emission of

319 pharmaceuticals to water bodies. In this study we use the proposed MAVT/MAUT

320 procedure to support the decision whether a wastewater treatment facility should be

321 installed in the hospital (and if yes which type) or if other measures to reduce

322 pharmaceuticals should be introduced. This is a good example of a complex decision

323 with conflicting objectives and different stakeholders. It also exemplifies a policy decision

324 under high uncertainty. Despite much research, scientifically sound cause-effect

325 relationships of micropollutants on aquatic organisms and ecosystems (Ankley et al.,

326 2007) are still lacking. A prominent exception is the negative effect of estrogens on fish

327 reproduction (Kidd et al., 2007). Notwithstanding this uncertainty and although there is

328 no legal directive for hospital wastewater treatment in Switzerland (yet), the directorates

329 of some hospitals are concerned. We chose the cantonal hospital Baden because it is

330 typical for a medium-sized, regionally important Swiss hospital, and because an on-site

331 pilot plant was installed to investigate different treatment methods from an engineering

332 point of view (McArdell et al., 2010). The directorate and legal authorities were very

333 interested in a comprehensive evaluation of these and other measures.

334   Details of the case study are given in Lienert et al. (2011), including an extensive

335 description of the decision alternatives, the prediction of consequences, and a

336 comparison of the results to those from a typical psychiatric clinic. This work is based on

337 an intensive collaboration between chemists, ecotoxicologists, wastewater engineers,

338 and decision analysts. The predictions concerning ecotoxicological effects and mass

339 flows of the hospital pharmaceuticals have been published in Escher et al. (2011). Here,

340 we focus on the decision support methodology (as conceptually outlined in section 2),

341 the preference elicitation, and on methodological implications on the results.

342   **3.2 Organizational steps of the project**

343   The project was carried out over a period of two years from 2008 to 2010. The whole

344 decision support procedure was divided into the following organizational steps (with

345 partially overlapping time frames):

346 1: Gather background information; obtain affirmation of the known most important

347   stakeholders to participate (contributed to A and B, Fig. 1).

348 2: First interview series (two to three hours each) for the stakeholder analysis and a

349   discussion of objectives, attributes, and some principal alternatives with selected

350   stakeholders. As an example, alternatives affecting nurses and patients were

351   discussed with hospital-internal actors (contributed to B, C, D, Fig. 1). To identify

352 stakeholders for step 4, we asked the interviewees to name all people that are

353 important for the decision problem or that could be affected by it. They then gave

354 numbers from 0 (not at all important/affected) to 10 (extremely important/strongly

355 affected by decision). The highest-ranked stakeholders on both axes were included in

356 the second interview series in step 4.

357 3: Data collection about alternatives and prediction of consequences (levels of attributes)

358 (contributed to D and E, Fig. 1).

359 4: Second interview series (for practical reasons restricted to about two hours each) with

360 representatives of the most important stakeholder groups identified in step 2 to elicit

361 their value function (F).

362 5: Calculate valuation and ranking of alternatives, sensitivity analyses with different

363 scenarios regarding weights, the form of value functions, probability distributions of

364 outcomes, and intrinsic risk attitudes of stakeholders (G).

365 6: Stakeholder workshop to discuss results (H).

366 7: Post processing, summarize results, draw conclusions.

367 We calculated values and utilities using the statistics and graphics software R

368 (http://www.r-project.org). Monte Carlo simulation (sample size 1000) was used to

369 propagate probability distributions of attributes to value and utility distributions and to

370 calculate expected utilities. In the next section we describe step 4 in more detail.

371 **3.3 Overview of preference elicitation for this case study**

372 As preparation, we sent a letter explaining the decision situation to each interview

373 partner. It included some information regarding the current situation (status quo), the

374 boundary conditions, and a description of the objectives and attributes. The amount of

375 information given to each main objective was similar (roughly same number of words) to

376 avoid the splitting bias.

377 All interviews were tape-recorded. At the beginning of the interview, we ensured the

378 comprehension of the information given in the letter and the agreement with it. We

379 invited interviewees to interrupt if anything was unclear and to express their thoughts.

380 This helped to reconstruct the arguments afterwards and to check the consistency of

381 answers. We emphasized that there were no right or wrong answers but that we were

382 interested in the opinion of the interviewee in representation of his or her stakeholder

383 group. We encouraged the interviewee to modify the answers until he or she was

384 confident that his or her opinion was reflected well.

3.4 Eliciting value functions for this case study

We elicited value functions instead of utility functions for the reasons mentioned in section 2.1: The elicitation of value functions is much easier and less prone to cognitive biases than the elicitation of utilities. The latter involves comparing lotteries instead of certain outcomes, which is usually demanding for the respondent. Secondly, we were interested in the strength of preference. Utilities however, only render the combination of strength of preference and risk-attitude. Finally, we wanted to assess the influence of prediction uncertainty on the valuation of outcomes and not only derive a ranking of the alternatives. This is important in cases where the analysis should facilitate a compromise between stakeholders and is not just intended to provide a final solution ("best alternative") to a decision problem. We evaluated the influence of different risk-attitudes on the ranking of alternatives by sensitivity analysis.

3.5 Aggregation of single-attribute value functions for this case study

We assumed an additive aggregation (Eq. 1). This could be justified by the nature of the objectives which were developed in the first round of interviews by considering the property of preferential independence. Moreover, we checked this assumption during the interview with trade-off questions (step 6 below). If stakeholders are able to specify their preferences with respect to the attribute levels of a subset of objectives independently of the attribute levels of the remaining objectives, preferential independence is indicated and an additive multi-attribute value function can be used (Eisenführ et al. 2010).
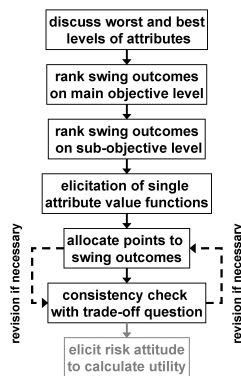
Fig. 2: Structure of the proposed elicitation process

The structuring of the interview is illustrated in Fig. 2, described below. The last step (in grey) was omitted and replaced by sensitivity analysis.

409

1. First, we presented the main objectives at the top hierarchical level (Table 1) and the worst and best level of the corresponding attributes. We included relevant background information (e.g., budget of hospital for objective "low costs"). This helps to take the range of attributes explicitly into account and relate it to the current situation and boundary conditions.

2. The interviewee then ranked hypothetical outcomes following the Reversed Swing method described above, where the attributes of only one main objective were at the worst, and the attributes of all others were at the best level. We did not yet show the sub-objectives to avoid the splitting bias; every main objective received equal attention. This procedure was facilitated by a tool we developed for this study, which illustrated the different outcome combinations, e.g., with emoticons and labels of the attribute levels. This tool is shown in the appendix (p. 2).

3. Afterwards, the ranking of sub-objectives was elicited in the same way.

4. Value functions for lowest level sub-objectives were elicited using the Midvalue Splitting method with consistency checks (all these depended on only one attribute; therefore we call them single-attribute value functions in the following). Due to time restrictions, we decided to elicit value functions at least for the attributes of the most important main objective by asking for the midvalues of the intervals [v=0,v=1], [v=0,v=0.5], and [v=0.5,v=1]. The midvalue of the interval [v=0.25,v=0.75] was used as consistency check. If disagreement between this point and the midvalue of the interval [v=0,v=1] occurred, the procedure was resumed. If necessary, further intervals were elicited. It was possible to resolve inconsistencies in all cases. Elicitation was facilitated by a graphical tool, a ruler labeled with the attribute range. The elicited midpoints were marked on this ruler. Additionally, the points were transcribed to a coordinate plane and interpolated to a value function for graphical examination by the interviewer.

   The value functions of the minor sub-objectives were elicited only with one point (mid-value of the interval [v=0,v=1]) or assumed to be linear. Sensitivity checks were performed a posteriori.

5. Reversed Swing: The stakeholders had to allocate points between 0 and 100 to the combination of hypothetical outcomes for all objectives, again attribute ranges were emphasized (see tool shown in the appendix, p. 2).

442 6. We then carried out consistency checks with Trade-Off questions for the two most

443   important main objectives: We confronted the interviewee with two hypothetical

444   outcomes that should be equally good (or bad) according to the elicited weights and

445   value functions. We asked for indifference between the two outcomes. If one was

446   preferred, the Swing weights were revised and the Trade-Off check repeated until the

447   interviewee was indifferent. The Trade-Off questions were also used to check for

448   preferential independence of the objectives as described above. (The tool for this step

449   is shown in the appendix, p. 5.)

450 7. Due to the reasons mentioned above, we did not elicit the risk attitude of the

451   interviewees but performed a sensitivity analysis a posteriori with two assumptions:

452   (1) all stakeholders are intrinsic risk neutral, i.e., multiattribute utility and value

453   functions are identical. (2) All stakeholders are intrinsic risk averse: the multi-attribute

454   value function is converted to a utility function by exponential transformation,

455   assuming an Arrow Pratt measure of absolute risk aversion of 4 (Arrow, 1965;

456   Menezes and Hanson, 1970; Pratt, 1964). A risk-seeking attitude was assumed

457   unrealistic in this application.

4. Results of the case study

459   The main results of the decision support process are presented below. For more

460 technical details and the comparison between two types of hospitals see Lienert et al.,

461 (2011).

4.1 Stakeholder analysis (B)

463   13 stakeholders were identified in phase 2 representing federal and cantonal

464 authorities (environmental, wastewater, and health authorities), hospital management,

465 physicians including hospital pharmacy, nursing staff, hospital technical services, and

466 experts (from ecotoxicology, environmental chemistry, and engineering).

4.2 Objectives hierarchy and attributes (C)

468   Four main and nine sub-objectives were identified in phase 2 (Table 1). Since

469 stakeholders could assign zero weights to sub-objectives they found irrelevant, we used

470 the same objectives hierarchy for all stakeholders. This facilitates preparing interviews

471 and comparing results between stakeholders.

472

473   Table 1: Objectives and attributes (WW = Wastewater, WWTP = Wastewater Treatment Plant)

Main objective	Sub-objective (*Abbreviation*)	Attribute	Unit	Range
Low costs	Low annual costs (***costs***)	Annual capital costs (investment and running costs)	CHF/year	0–1,500,000
Good wastewater quality	Low ecotoxicological risk potential (***ecotox***)	Ecotoxicological risk quotient (RQ) from the hospital WW arriving at the WWTP	RQ	0–10
	Low load of pharmaceuticals (***load***)	Load of pharmaceuticals in the hospital WW	kg/year	0–1400
	Low load of pathogens and (multi-)antibiotic-resistant bacteria (***bac***)	Relative load (compared to original amount) in WW	%	0–100
High feasibility	Low effort for nursing staff (***staff***)	Total mean additional workload of the whole nursing staff not accounted for in the costs	hours/day	0–6
	Low effort for patients (***patients***)	Fraction of patients in hospital unhappy with the measure	%	0–33
Good public perception	High positive media coverage (***posmed***)	Number of positive or neutral media articles in first half year	number	0–6
	Low negative media coverage (***negmed***)	Number of negative media articles in first half year	number	0–6
	High acceptance by authorities (***legal***)	Overall assessment of the particular alternative by authorities (MAVT result)	value	0–1

474

475   The four main objectives identified in this case study were low costs, good wastewater

476   quality, high feasibility, and good public perception. These main objectives were further

477   subdivided into sub-objectives as given in the second column of Table 1. For each sub-

478   objective, a measurable attribute had to be defined which measures the fulfillment of the

479   respective sub-objective in each alternative. In cases where natural attributes were not

480   available, we constructed attributes. For instance, as attribute for "high acceptance by

481   authorities", we calculated the distribution of overall values of each alternative for the

482   cantonal and federal authorities. We then averaged the value distributions over these

483   stakeholders (linear pooling) and used them as predictions for the non-authority

484   stakeholders. For the authority-stakeholders this sub-objective was omitted. To measure

485   the ecotoxicological risk potential of the wastewater, the risk quotient (RQ) was

486   estimated as described in Escher et al. (2011). The RQ is defined as predicted

487   environmental concentration (PEC) divided by the predicted no-effect concentration

488   (PNEC) and summed up over the mixture of pharmaceuticals.

489 For each of the alternatives described in section 4.3 below, the attributes had to be

490 predicted as described in section 4.4 below. The range defined in the last row of Table 1

491 covers the attribute range of the alternatives. As example, the range for the "fraction of

492 patients in hospital unhappy with the measure" is 0 – 33% because there is no

493 alternative which affects more than 33% of the patients in such a way that they could be

494 unhappy with the measure. The details, how we arrived at these case study-specific

495 estimates, are given in Lienert et al. (2011).

4.3. Alternatives (D)

497 To reduce pharmaceuticals, we considered on-site treatment of all hospital wastewater

498 (Total WW) by the following options for micropollutant removal: ozonation (O3),

499 powdered or granulated activated carbon, (PAC, GAC), reverse osmosis (RO), or

500 combinations of these. Furthermore, we considered the collection and treatment or

501 collection, transport, and incineration of the following source separated streams, which

502 contain a certain amount of the pharmaceutical load: wastewater collected with vacuum

503 toilets (Vacuum WC), urine where it is collected anyway (Urine) (i.e., catheters, bedpans,

504 urine bottles), urine collected with NoMix-technology (Urine NoMix), and urine collected

505 from patients that received X-ray contrast media with three pocket urinals named

506 roadbags (Roadbags). We distinguished the case that only stationary patients use

507 roadbags (Hospital) or stationary and out-patients use roadbags (Hosp+Outpatients).

508 The toilet wastewater contains (nearly) the full pharmaceutical load; vacuum toilets

509 having the advantage of requiring minimal flushing water, thus reducing storage and

510 transport volume. NoMix toilets collect the urine separately from feces and flush water

511 (see www.novaquatis.eawag.ch). Separate collection of urine could decrease 60–70% of

512 the pharmaceutical load (Lienert et al., 2007).

513 Additionally, the status quo (doing nothing) was included. The combination of different

514 collection pathways and technical pre- and post-treatment steps resulted in 68 viable

515 alternatives. Since many were similar, we grouped them and show here only the most

516 interesting selection.

4.4 Prediction of outcomes (E)

518 For each alternative and each attribute we derived a probability distribution from

519 experts (e.g., engineers, environmental chemists, ecotoxicologists) as described in

520 Lienert et al. (2011). Since the large number of alternatives resulted from combining

521 different collection and treatment steps, the number of expert predictions was

522 substantially smaller than the number of alternatives times the number of attributes.

523 Predictions concerning the actual load and ecotoxicological risk potential of the

524 pharmaceuticals are discussed in detail in the related study by Escher et al. (2011). The

525 uncertainty of the predictions was propagated to the results (values/utilities) with Monte

526 Carlo simulations. When only computing the expected utilities and aggregating them, as

527 is done in standard applications, the uncertainty in results cannot be evaluated.

528 4.5 Elicitation of subjective preferences (F)

529 The appropriateness of an additive aggregation was confirmed by all stakeholders in

530 the last step of the elicitation procedure, where we asked trade-off questions to check

531 the consistency of the weights assigned by the Reversed Swing procedure. All

532 stakeholders were able to specify their preferences with respect to the attribute levels of

533 two of the main objectives, independently of the attribute levels of the remaining

534 objectives, confirming preferential independence.

535 In Fig. A1(appendix p.6), the single-attribute value functions of the hospital director of

536 staff are shown as an example. The majority of the elicited value functions were non-

537 linear. This may be partly explained by the fairly large ranges of the attributes. Because

538 the final set of alternatives was not fully determined at this stage and the final results of

539 the predictions were not yet available (step 4 started before step 3 was completed), we

540 chose relatively large ranges to ensure that the predictions for all alternatives would be

541 included. The weights for sub-objectives of all stakeholders are shown in Fig. A2

542 (appendix p.7). The weighting was different for stakeholders, regardless to which

543 stakeholder group they belonged (authorities, hospital-internal actors, or experts).

544 4.6 Rank alternatives, analyze results (G)

545 Fig. 3 shows the expected value of the overall objective of each management option

546 (= alternative) for the director of staff. Additionally, we see the contribution of the sub-

547 objectives to the overall performance. The error bars show the 2.5 and 97.5% quantiles

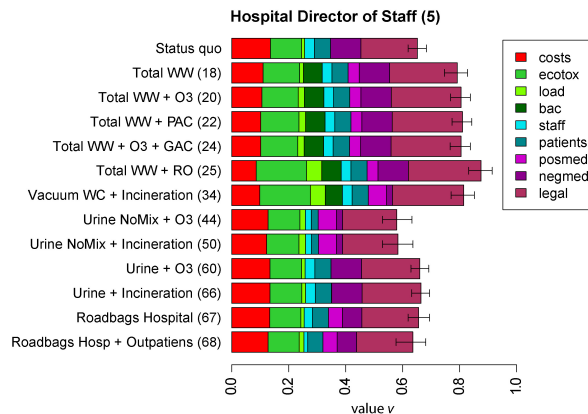548 of the value distributions from the Monte Carlo simulation.

Hospital Director of Staff (5)

Legend: costs, ecotox, load, bac, staff, patients, posmed, negmed, legal

Alternatives (top to bottom): Status quo, Total WW (18), Total WW + O3 (20), Total WW + PAC (22), Total WW + O3 + GAC (24), Total WW + RO (25), Vacuum WC + Incineration (34), Urine NoMix + O3 (44), Urine NoMix + Incineration (50), Urine + O3 (60), Urine + Incineration (66), Roadbags Hospital (67), Roadbags Hosp + Outpatiens (68)

x-axis: value *v*

Fig. 3: Selected alternatives for the hospital director of staff, x-axis shows the expected value of the overall objective, v, colored segments correspond to $w_i v_i$ (Eq. 1), error bars show 2.5 and 97.5% quantiles of the value distribution, for explanation of the legend see Table 1, names of alternatives are explained in the text in section 4.3.

This stakeholder gave the highest weight to the sub-objectives of "good public perception" (posmed, negmed, legal; Fig. A2). However, the objectives that discriminate most between alternatives are those with lower weights, namely the sub-objectives of "good wastewater quality" (ecotox, load, bac; Fig. A2). This can easily be seen in Fig. 3 when comparing the length of the colored bars. The largest differences between the alternatives exist in the green bars, the sub-objectives of "good waste water quality" (ecotox, load, bac), while the red and purple bars for "good public perception" are more or less of equal length. For this stakeholder, "Total WW + RO (25)" performs best, i.e., the total treatment of all wastewater with reverse osmosis (RO). RO virtually removes all micropollutants and pathogens, but is also the most expensive. Because several other alternatives have a huge overlap in the value distribution, it is not really possible to discriminate between them (alternatives 18, 20, 22, 24). These alternatives also treat the total wastewater and remove (nearly) all pharmaceuticals, depending on the method (O3; PAC, GAC). A similar performance is achieved by alternative 34 (vacuum toilets/incineration), which also removes all pharmaceuticals. The "Status quo" performs worse than these alternatives. Alternatives that treat only urine got even lower values by this stakeholder.

To compare the preferences of all stakeholders, the overall values and the ranking of the selected alternatives are given in Fig. 4. Confirming the result for the above stakeholder, also for most other stakeholders, the alternatives that collect and treat the whole wastewater stream on-site (or vacuum toilets/ incineration) perform quite well.

575 These alternatives dealing with the entire wastewater are depicted in blue and

576 systematically perform better than the "Status quo" (black line; Fig. 4). On the other hand,

577 those alternatives that only collect a certain fraction of the urine perform similar or even

578 worse for most stakeholders than the "Status quo"; i.e., the "NoMix"-alternatives (in red),

579 the alternatives collecting "urine where it is collected anyway" (orange) and the

580 "roadbags" (green). The high ranking of the "total wastewater treatment"-alternatives is

581 due to the predicted good performance of these alternatives regarding wastewater

582 quality, especially for the sub-objectives removal of (multi-)antibiotic resistant bacteria

583 and pathogens (bac) and ecotoxicological risk potential (ecotox). Ranked first for nearly

584 all stakeholders is Total WW + RO (25) (i.e., "reverse osmosis"), the most expensive

585 alternative (CHF 549'134 year^{-1}). That costs were not decisive might be due to the fact

586 that these costs are 0.3% of the annual budget of the hospital (Lienert et al., 2011).

587   However, Figure 4 also clearly shows that the discrimination *within* the general options

588 is small to negligible, especially if the uncertainty of the predictions is included in the

589 evaluation. For instance, for most stakeholders there is no difference between treating all

590 wastewater with "ozonation" (TotalWW+O3, alternative 20) or "powdered activated

591 carbon" (TotalWW+PAC, alternative 22); the lines on the value graph (left in Fig. 4)
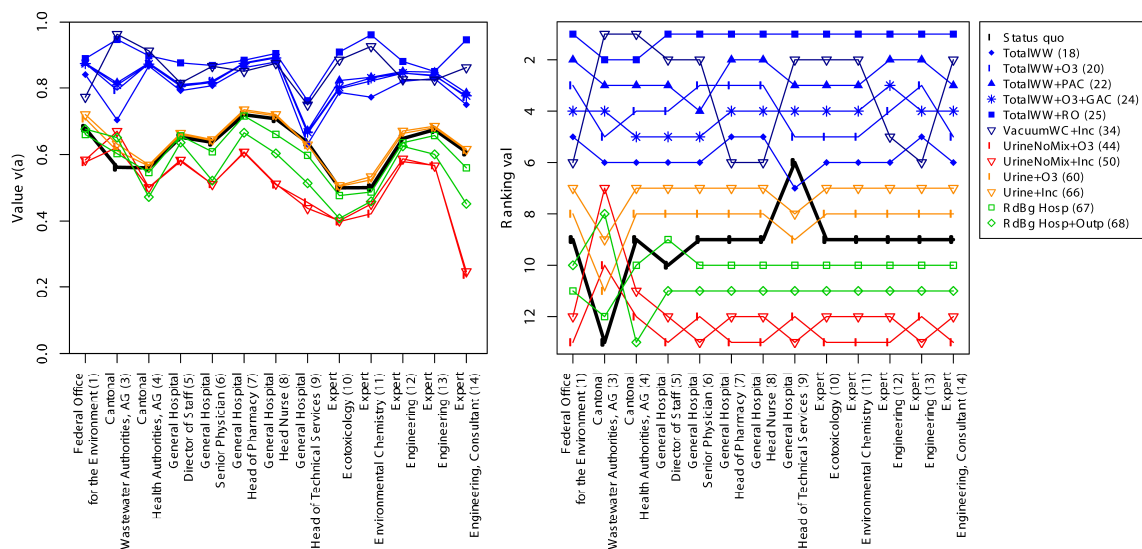
592 literally collapse.

Fig. 4: Expected value of the overall objective (left) and ranking (right) of alternatives for the different stakeholders. Abbreviations for alternatives are given in the text in section 4.3.

4.7 Sensitivity analyses (G)

596

597   Critical steps that were evaluated by sensitivity analyses are the elicited single-attribute

598   value functions and weights, the assumptions about intrinsic risk attitudes of

599   stakeholders, and the assessment of uncertainty of the predicted outcomes for the

600   wastewater quality attributes.

601   To test the robustness of the results regarding these assumptions we used the

602   following scenarios and compared their results with the original results:

603   **S1** Compare neutral (i.e., value = utility) and risk averse risk attitude

604   **S2** Double uncertainty (standard deviation) of ecotoxicological risk quotient (RQ;

605   ecotox) and load of pharmaceuticals (load), assuming risk aversion as in S1

606   **S3** Assume linearity for all single-attribute value functions

607   **S4** Assume non-linearity for all single-attribute value functions by replacing value

608   functions that were not elicited with value functions of other stakeholders

609   **S5** Increase weight of main objective "costs" by 25% (and re-scale all other

610   weights of main objectives accordingly)

611   **S6** Increase weight of main objective "wastewater quality" by 25% (and re-scale

612   all other weights of main objectives accordingly)

613   To assess the effects of these changed assumptions, one can assess the reversals in

614   the utility ranking or the change in overall expected values. If many alternatives are very

615   similar, rank reversals might occur even if changes in overall value are small. On the

616   other hand, major changes in the overall value might not lead to rank reversals if the

617   alternatives are very different or affected by the changes in a similar way. Therefore, it is

618   useful to look at both (Fig. 5, Table 2). The rank reversals were assessed by calculating

619   the Kendall correlation coefficient τ (Kendall, 1938) averaged over 13 stakeholders

620   (Table 2).

621

622 Table 2: Kendall correlation coefficients τ for the rankings of the six sensitivity scenarios S1 to S6 averaged

623 over 13 stakeholders ± standard deviation. We show results for the subset of alternatives presented in this

624 paper (n = 13) and the full set of alternatives (n = 68; Lienert et al., 2011). See text for explanations.

Scenarios	S1	S2	S3	S4	S5	S6
selected alternatives (n = 13)	1 ± 0	0.98 ± 0.03	0.95 ± 0.06	0.94 ± 0.06	0.97 ± 0.07	0.97 ± 0.02
all assessed alternatives (n = 68)	1 ± 0	0.97 ± 0.01	0.92 ± 0.08	0.90 ± 0.08	0.93 ± 0.09	0.95 ± 0.02

625

626 We observe a low number of rank reversals for the subset of alternatives (i.e., higher

627 correlation of SX with S1), but a slightly higher number for the whole set of alternatives.

628 This is not surprising since we did the selection because many alternatives of the original

629 set were very similar. Therefore, small changes in utility expectedly lead to much more

630 rank reversals in the original set than in the subset that discriminates strongly between

631 different alternatives. Below we discuss the sensitivity of the MCDA results to the

632 simplifying assumptions that facilitate the elicitation procedure.

633

reference, identical to Fig. 4a

S2 doubled stdev of ecotox and load

S3 assuming linearity for all single-attribute value functions

S4 assuming nonlinearity for all single objective value functions

S5 relative weight of costs +25%
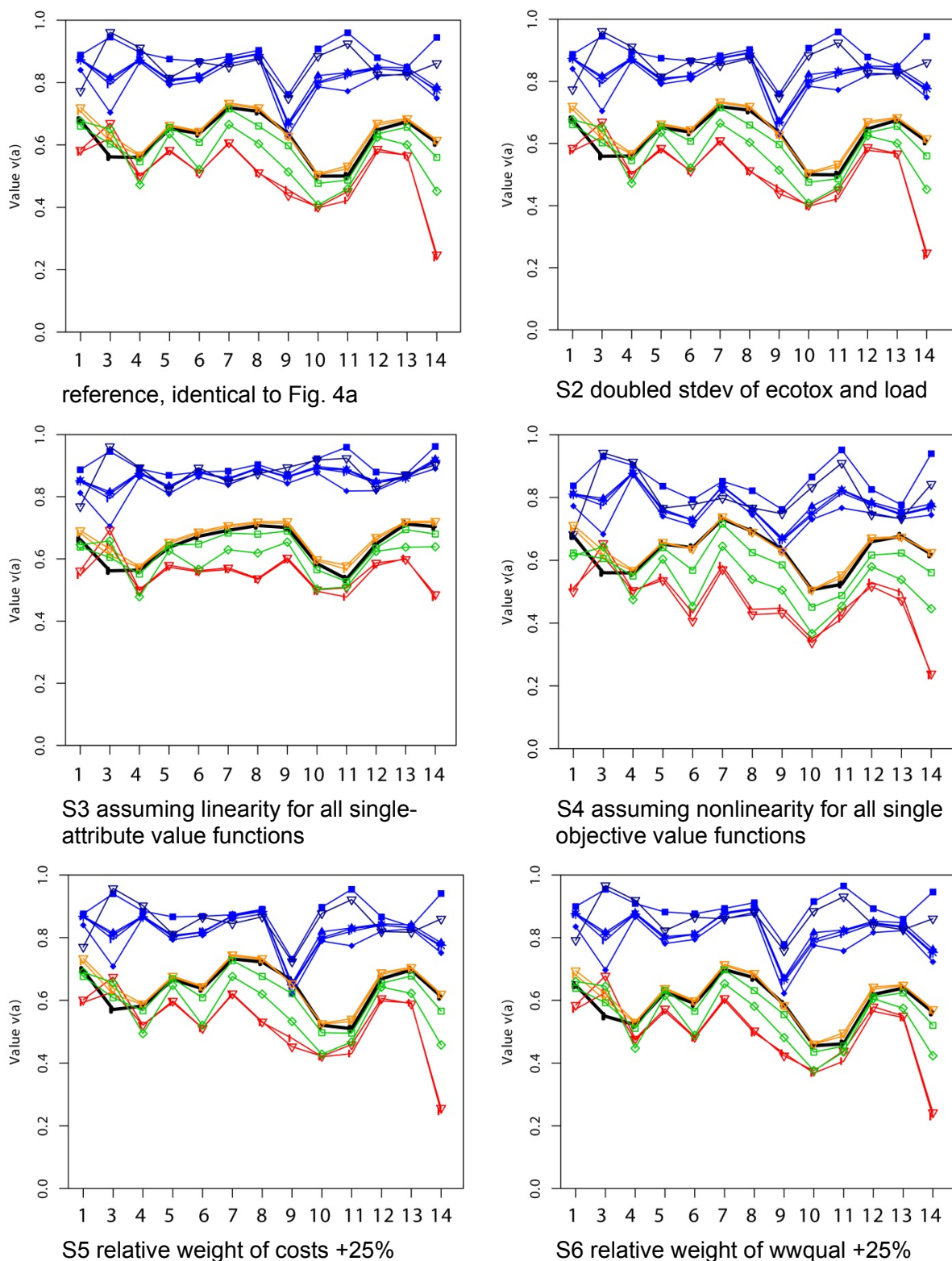
S6 relative weight of wwqual +25%

635   Fig. 5: Expected value of the overall objective of alternatives (y-axis) for the 13 different stakeholders (x-axis)

636   and different sensitivity scenarios (see text); legend and description of the stakeholders are given in Fig.4.

5. Discussion

5.1 Sensitivity of the results to the simplifying assumptions

Comparing the expected values (Fig. 5) shows that the general picture of all sensitivity scenarios is similar to the original: the blue alternatives (collection of all wastewater and on-site treatment, or transport/incineration) always have the highest values. Hence, the main results are robust and withstand changes in the underlying assumptions.

S1: Rank reversals when transforming values to utilities assuming risk aversion can be expected for alternatives that have nearly the same expected value but large differences in the uncertainty of the value distribution. This was not the case in our study. Our sensitivity analysis confirms that the risk attitude is not very influential in this decision (in the tested range covering risk neutrality to risk aversion with a Pratt Arrow Measure of 4; not shown in Fig. 5). This indicates that an elicitation of the risk-attitude of the stakeholders is not necessary to facilitate the decision. If this had not been the case, the risk attitude would have had to be elicited in an additional interview.

S2: Assuming a much higher uncertainty when predicting the wastewater quality attributes will only change the expected values in case of nonlinear value functions but in any case it might influence the ranking of expected utilities, if the risk attitude is not neutral. However, results show only minor changes in the ranking (Table 2) and in the expected values (Fig. 5).

S3: Assuming all single-attribute value functions to be linear leads to larger changes especially for stakeholder (9) (head of technical services) and to minor changes for others. This is due to highly nonlinear value functions for costs and the wastewater quality attributes of stakeholder (9). Therefore, assuming linear value functions is not always uncritical and should be validated in each application. However, the sensitivity test also shows that in our case uncertainties concerning elicited points of the value functions cannot have a high influence on major results.

S4: This scenario tests for the influence of our compromise due to time restrictions to elicit only the single-attribute value functions of the most important objectives. Assuming linearity for the value functions of the less important objectives has some influence on overall values and the ranking, but it would probably not lead to a different decision in our case study.

669 **S5** and **S6**: The rather high relative increase in weights for costs and wastewater quality
670   objectives did not lead to dramatic changes in the overall results, which were
671   generally quite robust to the weights. However, to discriminate similar alternatives,
672   the weights can play an important role.

5.2 Verification of the results in stakeholder workshop

674   We discussed the results with the stakeholders in a half-day workshop. They mainly
675 agreed with their own results of the MCDA. Furthermore, they agreed that the procedure
676 gave a lot of insight. For instance, they found it especially encouraging that such high
677 importance was placed on protecting the aquatic environment from pharmaceuticals by a
678 wide variety of stakeholders, including authorities, while low costs were not decisive in
679 this decision. The insights from this study will hopefully enter the discussion concerning
680 the reduction of micropollutants in water bodies. Currently, the Swiss Federal Office of
681 the Environment (www.bafu.ch) favors a centralized solution to upgrade large
682 wastewater treatment plants, but this (expensive) proposal has been challenged by
683 various stakeholder groups in a public hearing process. Our results clearly show that in
684 some cases cheaper point-source measures at hospitals might meet acceptance. In
685 addition, this may remain an option for hospitals that are not connected to a wastewater
686 treatment plant that will be upgraded. Finally, MCDA – although being quite laborious –
687 was judged to be helpful for structuring decisions and reaching consensus in other
688 environmental management situations as well.

5.3 Advantages of the proposed MCDA-procedure

690 We see the main benefits of such a careful MCDA process in the following:
691 1. The support of a fair discussion for group decisions or public decisions that involve
692   several stakeholder groups:
693   • due to the transparency of the procedure it is more difficult to influence results
694   with a hidden agenda than with unstructured negotiation processes;
695   • as the discussion focuses on valuing objectives rather than directly on
696   alternatives, the procedure facilitates appreciating other perspectives and opens
697   the horizon for new, creative alternatives;
698   • the objective prediction of consequences of the alternatives is explicitly separated
699   from the subjective valuation of consequences.

700 2. It is a convenient method for decisions with large sets of alternatives and attributes. The number of alternatives does not influence the elicitation procedure, only the effort in predicting consequences. A hierarchical structure of objectives facilitates handling many attributes.

704 3. It supports the creation of (new) consensus-alternatives since reasons for good or bad performance of alternatives are revealed, and new alternatives can easily be included in the analysis a posteriori.

707 A major challenge is the careful consideration of the choice and display of information during the decision making procedure (Dyer et al., 1992). This is an important point, because good information about the decision facts are necessary and attributes must be comprehensible to all stakeholders. On the other hand, stakeholders might be influenced by this information. Therefore, the amount and representation of prior information is a trade-off between providing sufficient information and exerting influence.

6. Conclusions

714 Using Multi-Attribute Value and Utility Theory (MAVT/MAUT) for policy support has the following conceptually important properties that are not all shared by other decision support techniques: (1) the preference function is independent of decision alternatives, this allows us to evaluate newly developed alternatives without redoing elicitation, (2) there are no methodological inconsistencies such as rank reversals when considering additional alternatives, (3) the aggregation procedure of values of sub-objectives to values of objectives at higher levels is (ideally) elicited from decision makers instead of applying a prescribed methodology, and (4) the combination of MAVT and MAUT allows us to consider the uncertainty of predictions and risk attitudes of decision makers. However, MAVT/MAUT has the disadvantage that the elicitation of value and utility functions can be very time-consuming and thus makes the technique difficult to apply. This study showed that a careful design that concentrates elicitation to sensitive branches of the value function can reduce the elicitation procedure of MAVT/MAUT to a limited time frame of two hours for each interview without losing much precision. However, the robustness of results to such shortcuts must be evaluated a posteriori carefully with sensitivity analyses. Adapting methods to elicit aggregation parameters with more realistic outcomes (Reverse Swing method) facilitated the interview task. A deliberate structure of the elicitation protocol helps avoiding common cognitive biases.

732   Explicitly considering uncertainty in the prediction of outcomes and propagating this
733   uncertainty to the resulting values is very important to assess the differences in the
734   evaluation of alternatives. Uncertainty in values clearly shows the significance of
735   differences in alternatives and thus indicates if several alternatives perform almost
736   equally well (see error bars in Fig. 3, and Fig. 4, left side). The calculation of expected
737   utilities alone does not provide this insight, but it makes it possible to consider the risk
738   attitude of the decision maker to rank alternatives.

739   Finally, we emphasize that this procedure is a tool for supporting decision-making by
740   stimulating a fair discussion between stakeholders with different perspectives, by
741   identifying causes of disagreement, and by inspiring the search for even better
742   alternatives than those evaluated in a first round. It is thus much more than a procedure
743   for ranking given alternatives. Dyer et al., (1992) pointed out that the opportunities for
744   MCDA applications in the public sector are unlimited and that there is a strong need for
745   good documented cases. With this study we tried to contribute to this. We hope to have
746   illustrated the usefulness of the MAVT/MAUT method in real decision situations,
747   characterized by a wide range of stakeholders that have limited time. We also hope to
748   stimulate the use of decision support methods to deal with other complex public and
749   political decision problems.

761   Appendix. Supplementary information

762   Supplementary information associated with this article can be found in the online
763   version (doi).

764 References

765 Abbas, E. A. (2010). Constructing Multiattribute Utility Functions for Decision Analysis. In J. J. Hasenbein
766   (Ed.), *TutORials in Operations Research* (Vol. 7, pp. 62-98). Hanover, MD: INFORMS.

767 Allais, M. (1953). Behavior of the Rational Man before Risk - Criticism of American School Postulates and
768   Axioms. *Econometrica, 21*, 503-546.

769 Ananda, J. & Herath, G. (2009). A critical review of multi-criteria decision making methods with special
770   reference to forest management and planning. *Ecological Economics, 68*, 2535-2548.

771 Ankley, G. T., Brooks, B. W., Huggett, D. B. & Sumpter, J. P. (2007). Repeating history: Pharmaceuticals in
772   the environment. *Environmental Science & Technology, 41*, 8211-8217.

773 Arrow, K. J. (1965). *Aspects of the Theory of Risk Bearing*. Helsinki: The Academic Bookstore.

774 Bana e Costa, C. A. & Vansnick, J. C. (2008). A critical analysis of the eigenvalue method used to derive
775   priorities in AHP. *European Journal of Operational Research, 187*, 1422-1428.

776 Barron, F. H., Von Winterfeldt, D. & Fischer, G. W. (1984). Empirical and Theoretical Relationships between
777   Value and Utility-Functions. *Acta Psychologica, 56*, 233-244.

778 Beier, S., Koster, S., Veltmann, K., Schroder, H. F. & Pinnekamp, J. (2010). Treatment of hospital
779   wastewater effluent by nanofiltration and reverse osmosis. *Water Science and Technology, 61* (7),
780   1691-1698.

781 Belton, V. & Stewart, T. J. (2002). *Multiple Criteria Decision Analysis. An integrated approach.* Boston:
782   Kluwer Academic Publishers.

783 Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463-501.

784 Bleichrodt, H., Pinto, J. L. & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the
785   prescriptive use of expected utility. *Management Science, 47*, 1498-1514.

786 Borcherding, K., Eppel, T. & von Winterfeldt, D. (1991). Comparison of Weighting Judgements in
787   Multiattribute Utility Measurement. *Management Science, 37*, 1603-1619.

788 Brans, J. P., Vincke, P. & Mareschal, B. (1986). How to Select and How to Rank Projects - the Promethee
789   Method. *European Journal of Operational Research, 24*, 228-238.

790 Brans, J. P. & Vincke, P. H. (1985). A Preference Ranking Organization Method - (the Promethee Method
791   for Multiple Criteria Decision-Making). *Management Science, 31*, 647-656.

792 Clemen, R. T. (1996). *Making hard decisions: an introduction to decision analysis* (2 ed.). Pacific Grove:
793   Brooks/Cole Publishing Company.

794 Cobb, C. W. & Douglas, P. H. (1928). A Theory of Production. *American Economic Review, 18*, 139-165.

795 Corner, J. L. & Kirkwood, C. W. (1991). Decision Analysis Applications in the Operations Research
796   Literature, 1970-1989. *Operations Research, 39*, 206-219.

797 Corsair, H. J., Ruch, J. B., Zheng, P. Q., Hobbs, B. F. & Koonce, J. F. (2009). Multicriteria Decision Analysis
798   of Stream Restoration: Potential and Examples. *Group Decision and Negotiation, 18*, 387-417.

799 Duckstein, L., Treichel, W. & El Magnouni, S. (1994). Ranking ground-water management alternatives by
800   multicriterion analysis. *Journal of Water Resources Planning & Management - ASCE, 120*, 546-565.

801 Dyer, J. S. (1990). Remarks on the Analytic Hierarchy Process. *Management Science, 36*, 249-258.

802 Dyer, J. S., Fishburn, P. C., Steuer, R. E., Wallenius, J. & Zionts, S. (1992). Multiple Criteria Decision-
803   Making, Multiattribute Utility-Theory - the Next 10 Years. *Management Science, 38*, 645-654.

804 Dyer, J. S. & Sarin, R. K. (1979). Measurable Multiattribute Value Functions. *Operations Research, 27*, 810-
805   822.

806 Dyer, J. S. & Sarin, R. K. (1982). Relative Risk-Aversion. *Management Science, 28*, 875-886.

807 Eisenführ, F., Weber, M. & Langer, T. (2010). *Rational Decision Making*. Heidelberg: Springer.

808 Escher, B. I., Baumgartner, R., Koller, M., Treyer, K., Lienert, J. & McArdell, C. S. (2011). Environmental
809   toxicology and risk assessment of pharmaceuticals from hospital wastewater. *Water Research, 45*, 75-
810   92.

811 Fishburn, P. C. (1967). Methods of estimating additive utilities. *Management Science, 13*, 435-453.

812 Gregory, R. S. & Keeney, R. L. (2002). Making smarter environmental management decisions. *Journal of the
813   American Water Resources Association, 38*, 1601-1612.

814 Hamalainen, R. P. & Alaja, S. (2008). The threat of weighting biases in environmental decision analysis.
815   *Ecological Economics, 68*, 556-569.

816 Hayashida T., Nishizaki I. & Ueda Y. (2010). Multiattribute utility analysis for policy selection and financing
817   for the preservation of the forest. *European Journal of Operational Research*, 200, 833-843.

818 Janssen, R. (2001). On the use of multi-criteria analysis in environmental impact assessment in The
819   Netherlands. *Journal of Multi-Criteria Decision Analysis, 10*, 101-109.

820 Joss, A., Siegrist, H. & Ternes, T. A. (2008). Are we about to upgrade wastewater treatment for removing
821   organic micropollutants? *Water Science and Technology, 57* (2), 251-255.

822 Joubert, A., Stewart, T. J. & Eberhard, R. (2003). Evaluation of water supply augmentation and water
823   demand management options for the City of Cape Town. *Journal of Multi-Criteria Decision Analysis, 12*,
824   17-25.

825 Kahneman, D. & Tversky, A. (1979). Prospect Theory - Analysis of Decision under Risk. *Econometrica, 47*,
826   263-291.

827 Keefer, D. L., Kirkwood, C. W. & Corner, J. L. (2004). Perspective on Decision Analysis Applications, 1990-
828   2001. *DECISION ANALYSIS, 1*, 4-22.

829 Keeney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decisionmaking*. Cambridge, MA:
830   Harvard University Press.

831 Keeney, R. L. & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*.
832   Cambridge: Cambridge University Press.

833 Keeney, R. L. & Wood, E. F. (1977). Illustrative Example of Use of Multiattribute Utility Theory for Water-
834   Resource Planning. *Water Resources Research, 13*, 705-712.

835 Kendall, M. G. (1938). A NEW MEASURE OF RANK CORRELATION. *Biometrika, 30*, 81-93.

836 Kidd, K. A., Blanchfield, P. J., Mills, K. H., Palace, V. P., Evans, R. E., Lazorchak, J. M. & Flick, R. W. (2007).
837   Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National
838   Academy of Sciences of the United States of America, 104*, 8897-8901.

839 Kiker, G. A., Bridges, T. S., Varghese, A., Seager, T. P. & Linkov, I. (2005). Application of multicriteria
840   decision analysis in environmental decision making. *Integrated Environmental Assessment and
841   Management, 1*, 95-108.

842 Kolpin, D. W., Furlong, E. T., Meyer, M. T., Thurman, E. M., Zaugg, S. D., Barber, L. B. & Buxton, H. T.
843   (2002). Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999-
844   2000: A national reconnaissance. *Environmental Science & Technology, 36*, 1202-1211.

845 Krzysztofowicz, R. & Duckstein, L. (1980). Assessment Errors in Multiattribute Utility-Functions.
846   *Organizational Behavior and Human Performance, 26*, 326-348.

847 Lenz, K., Mahnik, S. N., Weissenbacher, N., Mader, R. M., Krenn, P., Hann, S., Koellensperger, G., Uhl, M.,
848   Knasmuller, S., Ferk, F., Bursch, W. & Fuerhacker, M. (2007). Monitoring, removal and risk assessment
849   of cytostatic drugs in hospital wastewater. *Water Science and Technology, 56* (12), 141-149.

850 Lienert, J., Bürki, T., Escher, B. I. (2007) Reducing micropollutants with source control: substance flow
851   analysis of 212 pharmaceuticals in feces and urine. *Water Science and Technology* 56 (5), 87-96.

852 Lienert, J., Koller, M., Konrad, J., McArdell, C. & Schuwirth, N. (2011). Multiple-criteria decision analysis
853   reveals high stakeholder preference to remove pharmaceuticals from hospital wastewater.
854   *Environmental Science & Technology,* 45 (9), 3848-3857.

855 Linkov, I., Satterstrom, F. K., Kiker, G., Seager, T. P., Bridges, T., Gardner, K. H., Rogers, S. H., Belluck, D.
856   A. & Meyer, A. (2006). Multicriteria decision analysis: A comprehensive decision approach for
857   management of contaminated sediments. *Risk Analysis, 26*, 61-78.

858 Mareschal, B., De Smet, Y. & Nemery, P. (2008). Rank Reversal in the PROMETHEE II Method: Some New
859   Results. In *Ieem: 2008 International Conference on Industrial Engineering and Engineering*
860   *Management, Vols 1-3* (pp. 959-963).

861 Martin, W. E., Wise Bender, H. & Shields, D. J. (2000). Stakeholder objectives for public lands: Rankings of
862   forest management alternatives. *Journal of Environmental Management, 58*, 21-32.

863 McArdell, C. S. Kovalova, L., Eugster, J., Hagenbuch, M., Wittmer, A. & Siegrist, H. (2010). Elimination of
864   pharmaceuticals from hospital wastewater in a pilot membrane bioreactor with PAC or ozone post-
865   treatment. In *SETAC Europe, 20th Annual Meeting, 23.-27.05.2010* (pp.
866   http://www.eawag.ch/organisation/abteilungen/uchem/schwerpunkte/projektuebersicht/projekt2/index_E
867   N). Seville, Spain.

868 Mendoza, G. A. & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A
869   critical review of methods and new modelling paradigms. *Forest Ecology and Management, 230*, 1-22.

870 Menezes, C. F. & Hanson, D. L. (1970). On the Theory of Risk Aversion. *International Economic Review, 11*,
871   481-487.

872 Perez, J., Jimeno, J. L. & Mokotoff, E. (2006). Another potential shortcoming of AHP. *TOP, 14*, 99-111.

873 Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica, 32*, 122-136.

874 R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation
875   for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

876 Raju, K. S. & Vasan, A. (2007). Multi attribute utility theory for irrigation system evaluation. *Water Resources*
877   *Management, 21*, 717-728.

878 Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Sporri, C., Tockner, K. & Truffer, B. (2007).
879   Concepts of decision support for river rehabilitation. *Environmental Modelling & Software, 22*, 188-201.

880 Ríos-Insua, S., Gallego, E., Jiménez, A. & Mateos, A. (2006). A Multi-Attribute Decision Support System for
881   selecting intervention strategies for radionuclide contaminated freshwater ecosystems. *Ecological*
882   *Modelling, 196*, 195-208.

883 Roy, B. (1996). *Multicriteria methodology for decision aiding*. Dordrecht: Kluwer Academic Publishers.

884 Saaty, T. L. (1980). *The analytic hierachy process - planning, priority setting, resource allocation*. New York:
885   McGraw-Hill Inc.

886 Sarin, R. K. (1982). Strength of Preference and Risky Choice. *Operations Research, 30*, 982-997.

887 Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

888 Stewart, T. J. (1995). Simplified approaches for multicriteria decision making under uncertainty. *Journal of*

889 *Multi-Criteria Decision Analysis, 4*, 246-258.

890 Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y. M. & Wang, Q. N. (1996). Multiattribute

891 utility function for a comprehensive health status classification system - Health Utilities Index Mark 2.

892 *Medical Care, 34*, 702-722.

893 Varian, H. (2010). *Intermediate Microeconomics: A Modern Approach* (8 ed.): W. W. Norton & Company.

894 von Neumann, J. & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.): Princeton

895 University Press.

896 von Nitzsch, R. & Weber, M. (1993). The Effect of Attribute Ranges on Weights in Multiattribute Utility

897 Measurements. *Management Science, 39*, 937-943.

898 von Winterfeldt, D. & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, England:

899 Cambridge University Press.

900 Wang, X. & Triantaphyllou, E. (2008). Ranking irregularities when evaluating alternatives by using some

901 ELECTRE methods. *Omega, 36*, 45-63.

902 Weber, M. & Borcherding, K. (1993). Behavioral influences on weight judgments in multiattribute decision

903 making. *European Journal of Operational Research, 67*, 1-12.

904 Weber, M., Eisenführ, F. & Von Winterfeldt, D. (1988). The Effects of Splitting Attributes on Weights in

905 Multiattribute Utility Measurement. *Management Science, 34*, 431-445.