



**SGH**

**COLLEGIUM OF ECONOMIC ANALYSIS  
WORKING PAPER SERIES**

Choosing from multiple alternatives  
in cost-effectiveness analysis with fuzzy  
willingness-to-pay/accept and uncertainty

Michał Jakubczyk

1           **Choosing from multiple alternatives in cost-effectiveness**  
2           **analysis with fuzzy willingness-to-pay/accept and uncertainty**

3    Michał Jakubczyk

4           Decision Analysis and Support Unit, Warsaw School of Economics, Poland

5   Tippie College of Business, The University of Iowa, USA

6   **Abstract**

7           Cost-effectiveness analysis of medical technologies requires valuing health,  
8           an uneasy task, as confirmed by variability of published estimates. Treating  
9           the willingness-to-pay/accept (WTP/WTA) as fuzzy seems an intuitive  
10           solution. Based on this premise, I construct a framework allowing to compare  
11           multiple health technologies using choice functions. The final choice  
12           must be crisp, so I discuss various defuzzification methods and show that  
13           using *indecisiveness point (IP)* for WTP/WTA (the value the decision maker  
14           equally approves/disapproves) has desirable properties: satisfying the independence  
15           of irrelevant alternatives and not treating the Likert scale as interval. I suggest  
16           three approaches to infer about IP with Likert-based surveys in random samples  
17           (hypothesis testing, Bayesian or frequentist estimation). No difference between  
18           IPs for WTP/WTA is found, and an explanation of the WTP-WTA disparity is provided.  
19           Estimating IP results in stochastic uncertainty, and I show how to conduct sensitivity  
20           analysis in the framework and what new insight is gained.

21           **Keywords:** Willingness-to-pay/accept; Fuzzy sets; Preference elicitation;  
22           Cost-effectiveness analysis; Sensitivity analysis

23           **JEL:** C44; C13; D81; D61

24           **Highlights:**

- 25    ● Willingness-to-pay/accept (WTP/WTA) for health should be modelled  
26    as fuzzy concepts
- 27    ● I propose how to use fuzzy WTP/WTA in cost-effectiveness analysis
- 28    ● I show three methods to estimate the fuzzy WTP/WTA with Likert-  
29    based surveys
- 30    ● The fuzzy framework explains away the WTP-WTA disparity
- 31    ● The proposed framework provides new insight in sensitivity analysis
- 32

# 1 Introduction

Cost-effectiveness analysis (CEA) of health technologies (HT) require valuing life: determining the willingness-to-pay (WTP) for a unit of health (e.g. a quality-adjusted life year, QALY). Determining WTP feels difficult and apparently is, noting the variability of published results. Bellavance et al (2009) reviewed the literature on value of statistical life (VSL) and found standard deviations (SDs), across and within countries, approximately equal to the means. Lindhjem et al (2011) conducted a review in environmental, health, or transport context and SDs (based on standard errors) were twice as large as the means, these differing between the categories (4 million in health and 9 million in environmental context, in 2005 US\$). Other reviews confirm this variability (e.g. Viscusi and Aldy, 2003), also within a single country (Hultkrantz and Svensson, 2012, in Sweden). The heterogeneity is partially explained by, e.g., country or year. Doucouliagos et al (2012) discussed these issues, but still estimating VSL accounting for heterogeneity yielded a wide 95% confidence interval: (34–2,693) thousand 2000 US\$.

The variability is less surprising, with the non-market nature of health: *health services*, not health, are bought. The relation between the two is unclear (available to specialists, with inherent statistical uncertainty, and other uncertainties, e.g. the efficacy vs effectiveness) and translating the observed propensity to buy into WTP may mislead (due to paying via reimbursement not out-of-pocket, inconvenience or fear impacting the purchase decisions, or misjudging the risks to be reduced, cf. Andersson, 2007). Hence, no past market experience can precisely tell how we value health. Non-market goods are also specific regarding the relation of WTP to willingness-to-accept (WTA—a compensation demanded for a unit of good). Horowitz and McConell (2002) found on average  $WTA/WTP \approx 10$  for non-market, health, and safety-related goods vs 2–3 for other types. Thus, valuing health presents some difficulties, and explaining the WTP-WTA disparity should directly refer to the non-market character.

In health technology assessment (HTA—a process supporting the decisions which technologies to finance from public resources) various approaches were used to set the WTP. The regulator may not present a threshold or deny using any (e.g. United Kingdom), or the threshold may be legally defined (e.g. Poland, 125,955 Polish Zlotys, PLN, per QALY,  $\text{€}1 \approx 4.4 \text{ PLN}$ ). The threshold may be a multiple of gross domestic product per capita (in Poland, see also World Health Organization, 2001; Tan-Torres Edejer et al, 2003) or the cost of a QALY for some benchmark medical procedure (Lee et al, 2009). The thresholds commonly referred to (e.g. US\$50,000) may also reflect the convenience of round numbers (Grosse, 2008; Neumann et al, 2014). Setting the threshold impacts real decisions, so the ethical component emerges: refusing a treatment due to cost of QALY being \$1 too large sounds inhumane, and repudiates the readiness to define a threshold.

1 Jakubczyk and Kamiński (2015), onwards J&K, suggested thinking about  
2 WTP/WTA in terms of fuzzy set theory, a mathematical approach to modelling  
3 imprecise perceptions (Zadeh, 1965). This represents the lack of market experi-  
4 ence and the resistance against a precise threshold. J&K's show, based on survey  
5 results, that also HTA experts indeed perceive WTP/WTA fuzzily. I follow this  
6 path, making here three major contributions. Firstly, J&K defined the fuzzy pref-  
7 erence relation between HTs, effectively working with pair-wise comparisons. In  
8 HTA the choice is often made from more than two options, and the relation may  
9 not be transitive or complete, making it difficult to use. I show how to define  
10 choice functions in the fuzzy context. I discuss three approaches and advocate a  
11 particular one. Secondly, the respondents surveyed by J&K should be treated as  
12 random sample. I present three statistical methods (hypothesis testing, Bayesian,  
13 and frequentist) to formally calculate the parameters of the fuzzy model (I apply  
14 them to the same survey). The results show there is no WTP-WTA disparity in  
15 the present context. Thirdly, estimating the parameters results in stochastic uncer-  
16 tainty. I show how to combine it with other types of uncertainty in the sensitivity  
17 analysis. The new insights, as compared to standard methods used in CEA, ap-  
18 peal to intuition: considering technologies involving larger and larger trade-offs  
19 (i.e. offering larger effects at larger cost) increases the uncertainty present in the  
20 model under lack of conviction towards the exact WTP/WTA value. The partial  
21 results how to use choice functions in the fuzzy context in CEA were presented by  
22 Jakubczyk (2016), and here it is largely evolved, as, i), the present model allows  
23 the technologies to reduce the effectiveness (when WTA is used); ii), the single  
24 choice function presented there is shown to have unfavourable properties and a  
25 different one is advocated; iii), the methods of estimating the parameters of the  
26 model are presented; iv), the present model accounts for uncertainty.

27 The current paper, trying to comprehensively describe how to introduce fuzzi-  
28 ness to CEA, covers various aspects: decision modelling, statistical estimation,  
29 and Monte Carlo sensitivity analysis. Hence, a short overview and rationale for  
30 the structure is due. In section 2 I set the stage, formally defining the fuzzy  
31 WTP/WTA and presenting the survey. Analysing the data at this point shows  
32 why Likert-based questions should be used in eliciting WTP/WTA, which in turn  
33 promotes choice functions not requiring an interval-scale interpretation. Then, in  
34 section 3, I introduce three choice functions that can be used to select among de-  
35 cision alternatives, and recommend one. Applying this choice function requires  
36 calculating only one parameter of the fuzzy WTP/WTA, and in section 4 I present  
37 possible methods. The proposed choice function along with estimation methods  
38 replace the fuzziness with stochastic uncertainty, and I show in section 5 how to  
39 account for this (and other types of) uncertainty in sensitivity analysis and what  
40 the properties of the proposed methods are. I summarize the findings and present  
41 some outlook in the final section. The proofs are gathered in the appendix.

## 1 2 Fuzzy willingness-to-pay/accept

### 2 2.1 Fuzzy preferences on cost-effectiveness plane

3 Throughout the paper we compare HTs using two criteria: effectiveness and cost,  
4 denoted by  $(e, c)$  (subscripts added if needed). If  $(e, c)$  is known and WTP is set  
5 (and equal to WTA), then we select HT maximizing net benefit:  $NB = WTP \times$   
6  $e - c$  (cf. Garber, 2000). In the present paper we focus on the situation when  
7 WTP/WTA are not known precisely, and this imprecision is not of stochastic nature.

8 J&K defined a fuzzy preference relation,  $\mu : \mathbb{R}^2 \rightarrow [0, 1]$ , that  $\mu(e, c)$ ,  $(e, c) \in$   
9  $\mathbb{R}^2$ , measures the conviction that HT given by  $(e^* + e, c^* + c)$  is at least as good  
10 as HT  $(e^*, c^*)$ , irrespectively of  $(e^*, c^*) \in \mathbb{R}^2$  (based on *shift invariance* axiom).  
11 We will refer to  $\mathbb{R}^2$  as a *cost-effectiveness (CE) plane*. J&K's axioms imply:  
12 1)  $\mu(e, c) = 1$  in the IV quadrant (of CE-plane) with axes and the origin; 2)  
13  $\mu(e, c) = 0$  in the II quadrant with axes, without the origin; 3)  $\mu(e, c)$  equal on rays  
14 stemming from (not containing) the origin, i.e.  $\mu(e, c) = \mu(\gamma \times e, \gamma \times c)$ ,  $\gamma > 0$ ;  
15 4)  $\mu(e, c)$  increasing with  $e$  and decreasing with  $c$ ; 5)  $\forall e : \mu(e, c) = 0 (= 1)$  for  $c$   
16 large (negative) enough (*criteria tradeability*).

17  $\mu(\cdot, \cdot)$  is fully characterized by its values for  $e = 1$  and  $e = -1$  (and vice versa),  
18 motivating a definition of *fuzzy WTP (fWTP)*: a fuzzy number with membership  
19 function  $\mu_{fWTP}(x) = \mu(1, x)$ ,  $x \geq 0$ , and *fuzzy WTA (fWTA)*: with membership  
20 function  $\mu_{fWTA}(x) = \mu(-1, -x)$ ,  $x \geq 0$ . Figure 1 illustrates  $\mu$ ,  $\mu_{fWTP}$ , and  $\mu_{fWTA}$   
21 (as pictured, the axioms still allow non-trivial membership function).

22 The model nicely describes the relation between two technologies, e.g. when  
23 comparing a *status quo*,  $(e_1, c_1)$ , with a challenger,  $(e_2, c_2)$ : we then analyse  
24  $\mu(e_2 - e_1, c_2 - c_1)$  to see how convinced the decision maker is towards a change  
25 (and J&K show how to do it under uncertainty). Problems arise when we compare  
26 three HTs:  $A = (e_1, c_1)$ ,  $B = (e_2, c_2)$ , and the *status quo*, say a null option,  $(0, 0)$ .  
27 It is unclear which  $\mu$  to consider:  $\mu(A)$ ,  $\mu(B)$ ,  $\mu(B - A)$ , or  $\mu(A - B)$ ? It may  
28 happen that  $\mu(A) = 1$  and  $\mu(B - A) = 0$ , still telling nothing about  $\mu(B)$ ; e.g. in  
29 Figure 1 consider  $A = (1, -1)$  and  $B - A = (1, 3)$ ,  $(1, 4)$  or  $(1, 5)$ . It is, thus, dif-  
30 ficult to refer to any form of transitivity. It may happen that  $\mu(A - B) = 0$  and  
31  $\mu(B - A) = 0$ , i.e. the relation needs not be complete (in Figure 1 for  $A = (1, 2)$ ,  
32  $B = (-1, -2)$ ). The goal of the decision maker is to make a choice, not to perform  
33 a set of pairwise comparisons; and deriving the choice from the results of, neces-  
34 sarily pairwise, fuzzy preference measurements is not operational. In section 3 I  
35 take J&K's model in another direction.

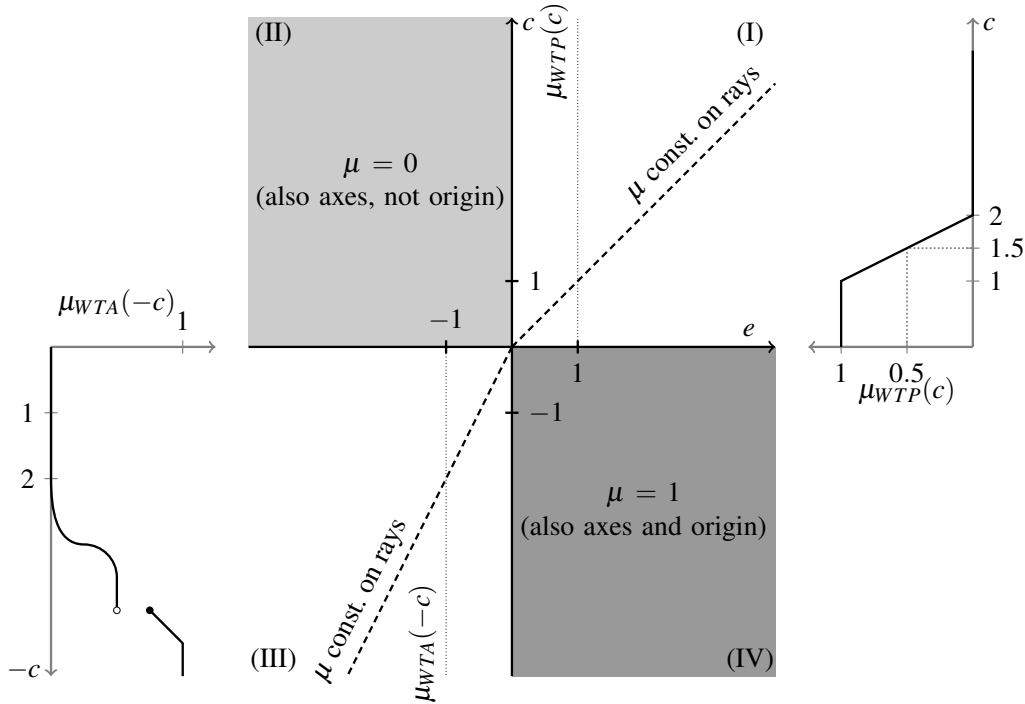


Figure 1: Fuzzy preference relation in cost-effectiveness plane (middle) and its relation to fuzzy willingness-to-pay/accept (right and left, respectively).

## 2.2 Survey results

Differently than J&K, fWTP/fWTA can be taken as the primitive of the model: instead of assuming that the decision maker has preferences for every  $(e, c)$ , we assume then that the decision maker has an (imprecise) idea of the value a unit of health (when gained or sacrificed) and accepts the axioms allowing to project it on the CE-plane (it suffices to assume that  $\mu_{WTP}(0) = 1$ ,  $\mu_{WTA}(0) = 0$ ,  $\mu_{WTP}(\cdot)$  is non-increasing,  $\mu_{WTA}(\cdot)$  is non decreasing, and both can be projected radially). It is then crucial to verify how decision maker perceives fWTP/fWTA and J&K surveyed HTA experts in Poland. This target group seems reasonable, being a proxy of an impersonal decision maker, while the general public may be unable to make an informed assessment (e.g. do not now the measures of effectiveness in HTA) and be biased by emotions (e.g. Johansson-Stenman and Svedsäter, 2012, showed that when valuing moral goods the respondents answer in a way that feels more socially-desirable). HTA experts are aware of the necessity to make trade-offs, so as to use public resources in the most efficient way. Nonetheless, the ideas presented in the present paper can be used with questionnaires collected in any group.

1 The details of the survey were presented by J&K. Among several questions 27  
2 respondents (5 were removed due to inconsistencies) were asked to assess their  
3 WTP and WTA, by reporting their conviction that a technology adding (sacrificing)  
4 one QALY should be used for a given cost increment (saving), for various  
5 cost differences (referred to as  $\lambda$ s for brevity, presented in Figure 2). The conviction  
6 was measured on Likert scale with five options: *definitely disagree*, *tend*  
7 *to disagree*, *I don't know*, *tend to agree*, *definitely agree*. From mathematical  
8 perspective it might be tempting to ask for a continuous  $[0, 1]$  valuation, but it  
9 is doubtful whether respondents can differentiate between the conviction, e.g. 0.8  
10 and 0.7, and what that would mean. Using a 5-option Likert is motivated, as levels  
11 can be assigned interpretation, e.g. *definitely agree* meaning *This is surely a good*  
12 *decision*, *tend to agree*—*I would make this decision, but clearly see downsides*, *I*  
13 *don't know*—*Can't tell if downsides or upsides are greater*, etc. The differences  
14 between the categories, alas, cannot be interpreted, which motivates building the  
15 framework based on the ordinal interpretation of the answers. The approach presented  
16 in subsequent sections would also work for a 3-level Likert.

17 Figure 2 (upper part for WTP, and the lower for WTA) presents the responses  
18 (vertical axis) for various  $\lambda$ s (horizontal axis, hundreds of 000s PLN/QALY). To  
19 no surprise, the individual experts differed, motivating the statistical approach to  
20 estimate the parameters of  $\mu_{\text{WTP}}$  and  $\mu_{\text{WTA}}$ . For option 3 individual respondents'  
21 answers are illustrated by horizontal bars spanning the  $\lambda$ s this option was selected  
22 for. For other options the area of the circles is proportional to the number of respondents.  
23 Black lines depict jumps across the middle answer (cf. section 4.2).  
24  $\mu_{\text{WTP}}(0) = 1$  is violated by one respondent selecting 4. This suggests that the respondent  
25 considered other aspects (e.g. allowed for the technology possibly causing  
26 adverse effects). This stresses the need to design questionnaires making the  
27 *ceteris paribus* condition maximally clear.

28 We may be tempted to check if  $\text{WTA} > \text{WTP}$ . This requires rephrasing the question  
29 in terms of fuzzy approach: we now ask if  $\mu_{\text{WTA}}$  is shifted rightward comparing  
30 to  $\mu_{\text{WTP}}$  (apart from a horizontal flip). The WTP-WTA disparity would then mean  
31 that  $\forall x \in \mathbb{R}_+ : \mu_{\text{WTP}}(x) \leq 1 - \mu_{\text{WTA}}(x)$ , and  $\exists x \in \mathbb{R}_+$  such that the inequality  
32 is strict (resembling the standard fuzzy numbers inequalities, Ramík and  
33 Římánek, 1985). It is not obvious how to conduct a statistical comparison (and  
34 still, with the survey we obtain Likert answers, not continuous membership). We  
35 might compare answers for fWTP vs fWTA (flipped around 3) using Wilcoxon  
36 paired test. If  $\text{fWTP} = \text{fWTA}$ , then  $H_0$  is true. Unfortunately, the test rejects  $H_0$   
37 even when fWTA is not shifted, e.g. when fWTA is flatter (options 2–4 used more  
38 often). Then, testing individual  $\lambda$ s separately would reject  $H_0$  in one direction for  
39 small  $\lambda$ s, and in the other for large, while the result for the pooled  $\lambda$ s depends on  
40 the structure of  $\lambda$  in the survey. A different approach is proposed in section 4.

41 The respondents were also asked to freely report their perceived WTP: a range

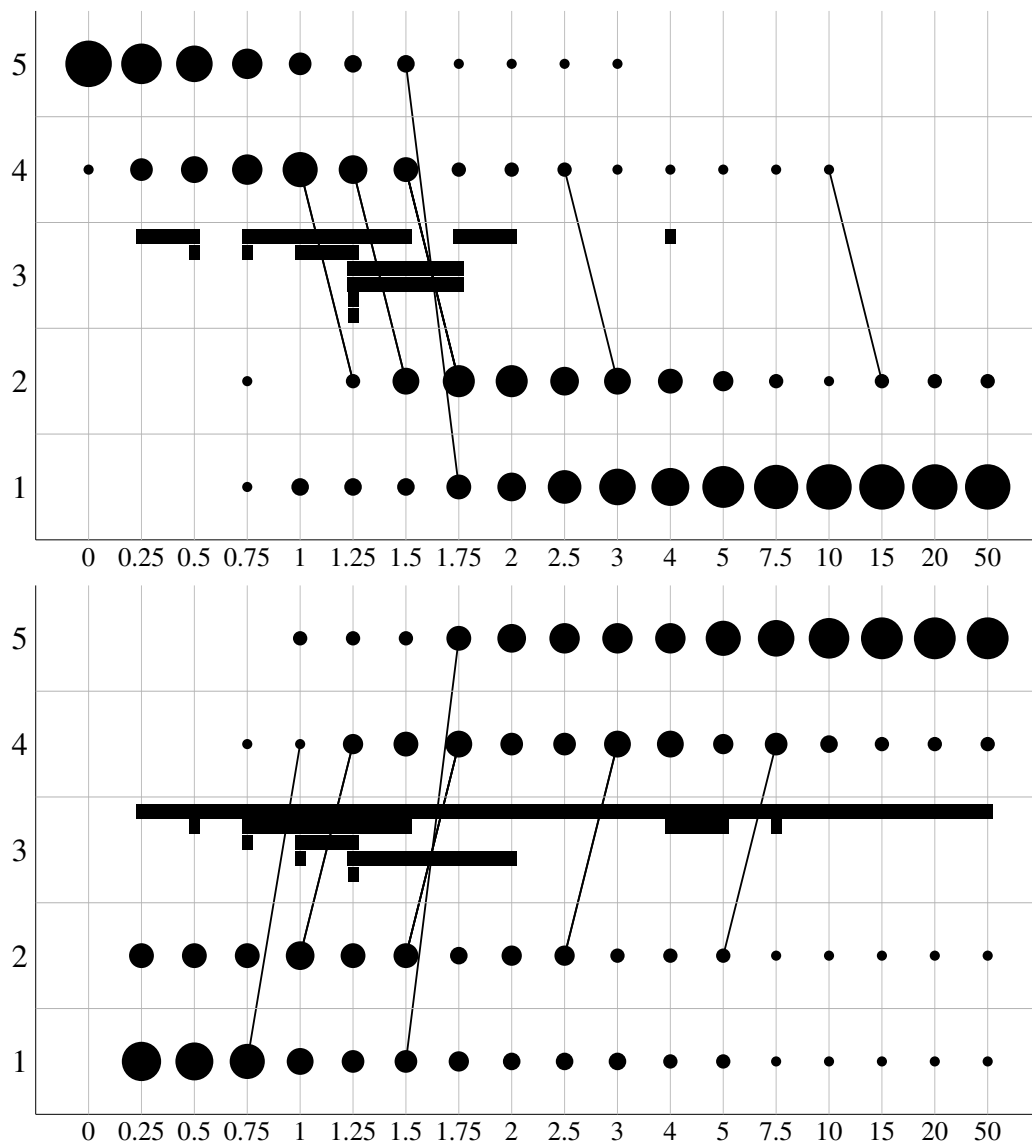


Figure 2: Survey results for WTP/WTA (above/below), values in horizontal axis in hundreds of 000s of PLN/QALY, answers (vertical axis) from a 5-level Likert scale (1—definitely disagree, 5—definitely agree). Horizontal bars represent individuals, circles—the fraction of respondents, lines—jumps across the middle option.

- 1 and a single value (unfortunately, this wasn't asked for WTA). On average the
- 2 range amounted to (88.9;125) and the single value to 105 (all results in 000s
- 3 PLN/QALY). Hence, the freely reported values corresponded to  $\lambda$ s towards which
- 4 the respondents felt quite convicted in Likert-based questions. For each respon-
- 5 dent I took the smallest range of  $\lambda$ s containing the whole freely reported range,



1 and calculated the average Likert response for these  $\lambda$ s (for a respondent reporting  
2 30–90 I consider  $\lambda = 25, 50, 75,$  and 100). The average (between the respondents)  
3 of these means amounted to 3.84, median to 4, 63.6% respondents had a mean  $\geq 4$ ,  
4 and only 2 respondents (9.1%)  $< 3$ . The analysis of Likert answers for the single  
5 freely reported WTP (if necessary, interpolating for two closest  $\lambda$ s) yields, simi-  
6 larly, the average of 3.92 and the median of 4. Thus, using Likert-based questions  
7 seems better at assessing the membership function for fWTP/fWTA than relying  
8 on directly reported ranges of values.

9 Averaging the freely reported values (105) and Likert answers (for individual  
10  $\lambda$ s) between the respondents, we find through interpolation that the single aver-  
11 age conviction towards the joint mean equals 3.63. Averaging the Likert answers  
12 for WTA, and proceeding backwards (assuming that the decision makers would  
13 also freely report WTA values towards which they feel convicted in Likert-based  
14 questions) yields the free value for WTA of 262.5. Thus, based on freely reported  
15 values, we would expect the WTP-WTA disparity of  $262.5 - 105 = 157.5$  (000s  
16 PLN/QALY) or a 2.5-fold difference. The above mechanism (of freely reporting val-  
17 ues still characterized by large conviction) will give rise to greater disparity, when  
18 fuzziness is large (i.e. respondents slower change their Likert response with  $\lambda$ s).  
19 This may explain why larger disparities are observed for non-market goods, when  
20 getting a crisp valuation is mentally more difficult.

## 21 **3 Fuzzy choice functions under certainty**

### 22 **3.1 Evaluating decision alternatives with fuzzy net benefit**

23 As mentioned in section 2, comparing alternatives with fuzzy preferences,  $\mu(\cdot, \cdot)$ ,  
24 may not be operational. Instead, we will now identify each option with a single  
25 fuzzy number—fuzzy net benefit (fNB), instead of two crisp numbers:  $c$  and  $e$ . I  
26 will then propose three methods how we can then compare these fuzzy numbers  
27 and select the greatest (defined in some way). Fuzzy NB represents the conviction  
28 of the decision maker that accepting a given option is equivalent to some monetary  
29 gain (the definition follows this of J&K).

30 **Definition 1** (fuzzy net benefit, fNB). *For any decision alternative, identified by*  
31  *$(e, c)$ , define fuzzy net benefit (fNB) as a fuzzy number with membership function*  
32  *$\mu_{fNB}$  given as  $\mu_{fNB}(x) = \mu(e, c + x)$ . I add  $(e, c)$  (or other symbol denoting the*  
33 *alternative) if needed to avoid confusion:  $fNB(e, c)$  and  $\mu_{fNB(e,c)}(x) = \mu(e, c + x)$ .*

34 As  $\mu$  is constant on rays, fNB can be equivalently defined using fWTP/fWTA  
35 (the notation simplifies further, taking  $\mu_{fWTP}(x), \mu_{fWTA}(x) = 1$  for  $x < 0$ ; this  
36 approach is used in the proof of Lemma 1):

- 1 • for  $e = 0$  we take  $\mu_{\text{fNB}(0,c)}(x) = \mathbf{1}_{(-\infty, -c]}(x)$ ,
- 2 • for  $e > 0$  we take  $\mu_{\text{fNB}(e,c)}(x) = \mu_{\text{fWTP}}(\max((c+x)/e, 0))$ ,
- 3 • for  $e < 0$  we take  $\mu_{\text{fNB}(e,c)}(x) = \mu_{\text{fWTA}}(\max((c+x)/e, 0))$ .

4 Effectiveness and cost for considered options are measured relative to *status*  
5 *quo*—the treatment that would be provided if no decision were made. The selec-  
6 tion of *status quo* is important; since we differentiate between WTP and WTA,  
7 changing the *status quo* may change if a given HT is effect-enhancing or reduc-  
8 ing, and so whether WTP or WTA are applied. That the selection of *status quo*  
9 may change the final decision motivates making the selection meaningful and  
10 representing the actual state of the world. Still, conveniently, having looked at  
11 the specific choice functions (subsection 3.2) and estimates of their parameters  
12 (section 4) we will see that with current dataset the selection of *status quo* is not  
13 important in the certainty case, as WTP-WTA disparity disappears. Another issue  
14 is, that the *status quo* may be a composite, i.e. a mix of treatments is currently  
15 used in patients. This will come back in section 5, when discussing uncertainty.

16 The interpretation of  $\text{fNB}(e, c)$  is the following: the decision maker agrees  
17 with conviction  $\mu_{\text{fNB}(e,c)}(x)$  that adopting HT characterized by  $(e, c)$  (relative to  
18 *status quo*) would be attractive (compared to *status quo*) even if it costed  $x$  more.  
19 In other words, she agrees with this conviction that adopting  $(e, c)$  is equivalent  
20 to gaining a monetarily-expressed pay-off of  $x$ . It thus is intuitive to compare deci-  
21 sion alternatives based on fNB. The shape of  $\mu_{\text{fNB}}$  is identical as the shape of  
22  $\mu_{\text{fWTP}}$  (or  $\mu_{\text{fWTA}}$ ), for  $\mu_{\text{fWTP}}$  from Figure 1 exemplary  $(e, c)$  values with corre-  
23 sponding fNBs are presented in Figure 3.

24 To strengthen the rationale for using fNB when comparing options, I verify  
25 how it behaves in obvious cases of dominance or (less obvious) extended domi-  
26 nance. This is easier done working on  $\alpha$ -cuts of fNB. An  $\alpha$ -cut of a fuzzy number  
27  $F$  defined on domain  $\mathbb{R}$  with membership function  $\mu_F(\cdot)$  will be denoted by  $A_F(\alpha)$   
28 and defined as

$$A_F(\alpha) = \{x \in \mathbb{R} : \mu_F(x) \geq \alpha\}, \quad (1)$$

29 for  $\alpha \in (0, 1]$  and  $A_F(0) = \cup_{\alpha \in (0, 1]} A_F(\alpha)$ . The following useful lemma holds (be-  
30 cause we work on *sets* we have to use special addition and product operators).

31 **Lemma 1.** *Take any  $\alpha \in (0, 1]$ .  $A_{\text{fNB}(e,c)}(\alpha)$  is linear with respect to  $(e, c)$ , where*  
32  *$c \in \mathbb{R}$ , and either  $e \geq 0$  or  $e \leq 0$ , in a sense that*

- 33 •  $A_{\text{fNB}(e_1+e_2, c_1+c_2)}(\alpha) = A_{\text{fNB}(e_1, c_1)}(\alpha) \oplus A_{\text{fNB}(e_2, c_2)}(\alpha)$ ,
- 34 •  $A_{\text{fNB}(\gamma e, \gamma c)}(\alpha) = \gamma \odot A_{\text{fNB}(e, c)}(\alpha)$  for any  $\gamma > 0$ ,

35 where  $A \oplus B := \{x + y : x \in A \wedge y \in B\}$  and  $\gamma \odot A := \{\gamma x : x \in A\}$ .

1 Two corollaries follow.

2 **Corollary 1.** Assume  $(e_2, c_2)$  is Pareto-dominated by  $(e_1, c_1)$ , i.e.  $e_2 \leq e_1 \wedge c_2 \geq c_1$   
3 (at least one inequality strict). Then  $\forall_{\alpha \in (0,1]} : A_{fNB(e_2, c_2)} \subset A_{fNB(e_1, c_1)}$ . Moreover,  
4 if  $c_2 > c_1$  or  $(e_2 < e_1$  and  $\mu_{fWTP}(x) > 0$ ,  $\mu_{fWTA}(x) > 0$  for some  $x > 0$ ), then  
5  $\exists_{\alpha \in (0,1]} : A_{fNB(e_2, c_2)} \neq A_{fNB(e_1, c_1)}$ .

6 Notice that  $(e_1, c_1)$  and  $(e_2, c_2)$  can be in any quadrants of CE-plane and that  
7 the implication can be seen as the most typical fuzzy numbers inequality (see, e.g.  
8 Ramík and Římánek, 1985). In the following corollary we consider points in a  
9 predetermined half of the plane.

10 **Corollary 2.** If  $(e_3, c_3)$  is extended dominated by  $(e_1, c_1)$  and  $(e_2, c_2)$ , i.e.  $(e_3, c_3)$   
11 is Pareto-dominated by some  $\gamma(e_1, c_1) + (1 - \gamma)(e_2, c_2)$  with  $\gamma \in [0, 1]$ , and either  
12  $e_1, e_2, e_3 \geq 0$  or  $\leq 0$ . Then  $\forall_{\alpha \in (0,1]} : A_{fNB(e_3, c_3)} \subset (A_{fNB(e_1, c_1)} \cup A_{fNB(e_2, c_2)})$ .

13 We cannot use the above corollary to infer about points in different halves of  
14 CE-plane: take  $(e_1, c_1) = (1, 1.5)$ ,  $(e_2, c_2) = (-1, -2)$ ,  $(e_3, c_3) = (0, 0)$ , and  $\mu$  as  
15 in Figure 1. Then  $(e_3, c_3)$  is extended dominated but, e.g. the 1-cut for  $fNB(e_3, c_3)$   
16 is not a respective subset.

17 The two corollaries confirm that fNB behaves intuitively and also may be used  
18 to quickly eliminate alternatives having no chances of being picked up by specific  
19 choice functions (as defined in the next subsection). Figure 3 illustrates exemplary  
20  $\mu_{fNB}$  and the corollaries in work.  $X$  is dominated by  $A$ , and respective  $\alpha$ -cuts are  
21 subsets (would be for  $\mu_{fWTP}$  shaped in any way), illustrated by membership function  
22 being shifted to the left. On the other hand, even though  $A$  is not dominated  
23 by  $B$ , its  $\alpha$ -cuts are subsets, but would cease to be for some other  $\mu_{fWTP}$ .  $Y$   
24 is extended dominated by  $B$  and  $C$ , and its  $\alpha$ -cuts are subsets of respective unions  
25 (not by  $\alpha$ -subsets of only  $B$  or  $C$ ).

## 26 3.2 Making a choice

27 In a standard, crisp approach in HTA the decision is made by maximizing the  
28 regular, crisp NB. *Per analogiam*, we want to make decision now by maximizing  
29 fNB, and below I present three possible approaches. We consider  $n$  alternatives,  
30 denoted by  $D_i$ , where  $i \in I = \{1, \dots, n\}$  and  $D_i = (e_i, c_i)$ . Each approach is a  
31 choice function, prescribing a crisp choice from a given menu of alternatives. It  
32 is easier to derive two of them when considering only HTs in quadrants I & IV of  
33 the CE-plane. These approaches can still be used for HTs in the whole CE-plane.  
34 Importantly none of the approaches violates the dominance (also extended), which  
35 easily follows from the above corollaries.

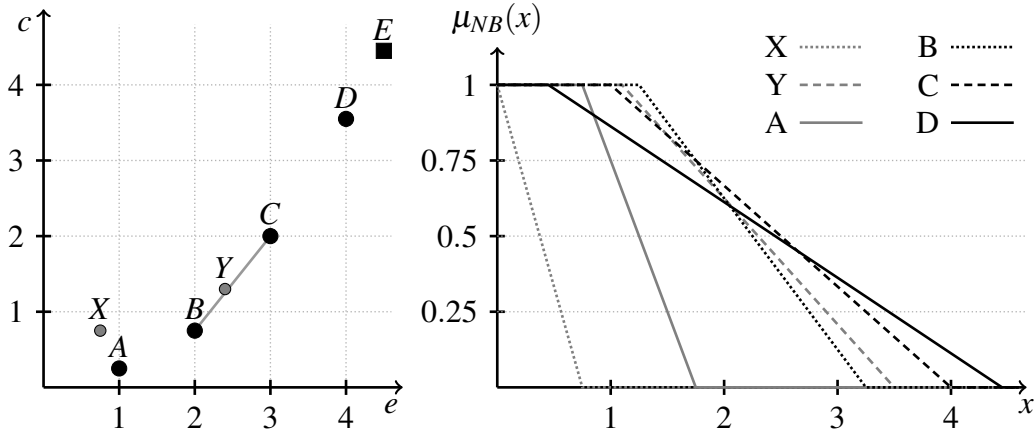


Figure 3: Exemplary fNB membership functions, when  $\mu_{WTP}$  decreases linearly in  $[1; 2]$  (as in Fig. 1).  $X = (0.75, 0.75)$ ,  $Y = (2.4, 1.3)$ ,  $A = (1, 0.25)$ ,  $B = (2, 0.75)$ ,  $C = (3, 2)$ ,  $D = (4, 3.55)$ , and  $E = (4.5, 4.45)$ , fNB for E not drawn for clarity.

### 1 3.2.1 Conviction of bestness

Start with  $n = 2$ , no Pareto-domination,  $e_1, e_2 \geq 0$ , and, without loss of generality,  $e_2 > e_1$ . Thinking in terms of linearity (Lemma 1), we can intuitively identify the conviction that  $D_2$  is not worse than  $D_1$  with the conviction that  $\text{fNB}(D_2 - D_1) \geq 0$ , hence,  $\mu_{\text{fNB}(D_2 - D_1)}(0)$ . Now, consider  $D_3$ ,  $e_3 > e_2$ . Analogously, the conviction that  $D_2$  is best equals the conviction that fNB (non-strictly) increases between  $D_1$  and  $D_2$ , and does not strictly increase between  $D_2$  and  $D_3$ . Mathematically, it is the conviction that  $\text{fNB}(D_2 - D_1) \geq 0$  and not  $\text{fNB}(D_3 - D_2) > 0$ , which now requires selecting the fuzzy logical operators (AND, NOT). I take NOT  $\text{fNB}(D_3 - D_2) > 0 = 1 - \mu_{\text{fNB}(D_3 - D_2)}(0)$  (typical approach), and the bounded sum AND (see, e.g. Smithson, 1987) to get

$$\max \{ \mu_{\text{fNB}(D_2 - D_1)}(0) + (1 - \mu_{\text{fNB}(D_3 - D_2)}(0)) - 1, 0 \}$$

- 2 or  $\max \{ \mu_{\text{fNB}(D_2 - D_1)}(0) - \mu_{\text{fNB}(D_3 - D_2)}(0), 0 \}$ , which equals 0 for  $D_2$  extended  
3 dominated by  $D_1$  and  $D_3$ . The above calculations can be represented graphically:  
4  $\mu_{\text{fNB}(D_2 - D_1)}(0)$  is the length of the segment of  $\alpha$ s that  $A_{\text{fNB}(D_1)}(\alpha) \subset A_{\text{fNB}(D_2)}(\alpha)$   
5 (and this representation motivates the selection of AND operator). Based on this  
6 reasoning, I assign each  $D_i$  the following conviction that it maximizes the fNB:

$$\beta_i := \int_0^1 \left[ \bigcup_{j \in I} A_{\text{fNB}(D_j)}(\alpha) \subset A_{\text{fNB}(D_i)}(\alpha) \right] d\alpha, \quad (2)$$

- 7 where  $[P] = 1$  if  $P$  is true, and 0 otherwise. Several options may have  $\beta > 0$ , as  
8 the decision maker is not fully confident of her WTP/WTA. Using this approach to

1 make a final decision it would be natural to select  $\arg \max_{i \in I} \beta_i$ . In the example in  
 2 Figure 3 we have  $\beta_X = \beta_Y = \beta_A = 0$ ,  $\beta_B = 0.25$ ,  $\beta_C = 0.3$ , and  $\beta_D = 0.45$  (ignore  
 3  $E$  for now). This approach is the most fuzzy one: it postpones the crispification  
 4 until the last possible moment, just when the crisp choice is being made.

5 More technically,  $\beta$ s can be calculated (i.e. integrals in equation 2 are well  
 6 defined): Lemma 1 and monotonicity of  $\mu_{\text{fNB}}$  (from the monotonicity of  $\mu_{\text{fWTP}}$ )  
 7 imply that the integrand will be equal to 1 over a single segment of  $\alpha$ s. The situ-  
 8 ation gets complicated when considering the whole CE-plane. No so intuitive  
 9 derivation can be presented (still, the method and its graphical interpretation ap-  
 10 peals to intuition). With no additional stipulations regarding  $\mu_{\text{fWTP}}$  and  $\mu_{\text{fWTA}}$  the  
 11 resulting  $\mu_{\text{fNB}}$  for various alternatives can intersect many times (countably many  
 12 at maximum, though) for various  $\alpha$ s, and so we would have to add up the lengths  
 13 of several segments in equation 2. This is unlikely to cause any problems in real  
 14 applications ( $\mu_{\text{fWTA}}$  and  $\mu_{\text{fWTA}}$  would be approximated by regular functions) and  
 15 is not pursued here.

16 There are at least two disadvantages to basing the choice on  $\beta$ . Firstly, we  
 17 need to interpret membership as the interval scale to calculate the vertical distance,  
 18 e.g. deriving  $\mu_{\text{fWTP}}$  and  $\mu_{\text{fWTA}}$  from Likert-based questions, we need to interpret  
 19 differences between consecutive options. Secondly, the basing the choice on  $\beta$ s  
 20 violates the Chernoff property (or Independence of Irrelevant Alternatives) of a  
 21 choice function (see, e.g. Sen, 1970). An example: consider also  $E$  in Figure 3.  
 22 The  $\mu_{\text{fNB}(E)}$  would be very flat, intersecting with  $\mu_{\text{fNB}(D)}$  for  $\alpha = 0.2$ . Now  $\beta_B =$   
 23  $0.25$ ,  $\beta_C = 0.3$ ,  $\beta_D = 0.25$ , and  $\beta_E = 0.2$ , and so  $C$  should be chosen (even though  
 24 available before and not recommended). Therefore I do not recommend using  
 25  $\beta$ s to drive decisions, but find them useful in sensitivity analysis, as presented in  
 26 section 5.

### 27 3.2.2 Average fuzzy net benefit

28 The remaining two choice functions are based on crispifying fNB, and comparing  
 29 these crisp representations. First, I continue with measuring the vertical changes  
 30 in  $\mu_{\text{fNB}}$ , but I interpret them as probabilities (conveniently summing to 1; the  
 31 membership function is treated as a, perhaps flipped horizontally, cumulative dis-  
 32 tribution function). Then for each  $D_i$ , take

$$\tau_i(\alpha) := \sup A_{\text{fNB}(D_i)}(\alpha), \quad (3)$$

33 and calculate an average fNB (ANB):  $\text{ANB}_i := \int_0^1 \tau_i(\alpha) d\alpha$ , i.e. average out the  
 34 bounds of  $\alpha$ -cuts. Then the choice is simply  $\arg \max_{i \in I} \text{ANB}_i$ . Technically, the  
 35 integral exists, as  $\tau$  is non-decreasing and bounded (for a given  $(e, c)$ ). Consider-  
 36 ing the complete CE-plane does not change the intuition behind the derivation nor

1 the feasibility to use. The method preserves the Chernoff property: the evaluation  
 2 of each alternative is independent of other options. The obvious disadvantage is,  
 3 again, the necessity to interpret the membership function as an interval scale.

### 4 3.2.3 Median fuzzy net benefit

5 A natural solution to avoid interval, and focus on ordinal, interpretation is to com-  
 6 pare medians, not means. Hence, the choice function I recommend in the current  
 7 framework is to maximize  $\tau_i(0.5)$  (eq. 4), i.e. the supremum of the 0.5-cut of fNB,  
 8 henceforth *median fNB (MNB)*, formally:

$$\text{MNB}_i := \sup A_{\text{fNB}(D_i)}(0.5) = \tau_i(0.5). \quad (4)$$

9 MNB can be interpreted as a value that the decision maker equally agrees/disagrees  
 10 that is a monetary equivalent of using a given technology. Maximizing MNB, as a  
 11 decision making rule, can always be applied (no fancy integrals) and preserves the  
 12 Chernoff property. It can also be used for the complete CE-plane, with the same  
 13 interpretation. In the example in Figure 3 MNB selects  $C$  (due to the piece-wise  
 14 linearity of the membership functions, maximizing ANB leads to the same choice,  
 15 but in general the outcomes would differ).

16 There are formal arguments motivating using MNB. As shown by Corollary 1  
 17 the dominated technologies are characterized by fNB included (via standard fuzzy  
 18 set inclusion) in some other fNB. In case of no dominance this inclusion may  
 19 not hold, but it can be shown that fNB of the MNB-maximizing option *weakly*  
 20 *includes* other fNB (using definition of Dubois and Prade, 1980).

**Proposition 1.** *Take  $n$  decision alternatives,  $D_i$ . If  $D_{i^*}$  maximizes MNB, then  $\text{fNB}_{i^*}$  weakly includes  $\text{fNB}_i$  for any  $i$  not maximizing MNB, i.e.<sup>1</sup>*

$$\inf_{x \in \mathbb{R}} \max (\mu_{\text{fNB}(D_{i^*})}(x), 1 - \mu_{\text{fNB}(D_i)}(x)) \geq \frac{1}{2},$$

and  $\text{fNB}_i$  weakly includes  $\text{fNB}_{i^*}$  at maximum to the same degree:

$$\inf_{x \in \mathbb{R}} \max (\mu_{\text{fNB}(D_i)}(x), 1 - \mu_{\text{fNB}(D_{i^*})}(x)) \leq \frac{1}{2}.$$

21 *Moreover, two implications hold:*

- 22 • *if  $\mu_{\text{fWTP}}$  and  $\mu_{\text{fWTA}}$  are strictly decreasing (for values within  $(0, 1)$  interval),*
- 23 *then the above inequalities are strict;*

---

<sup>1</sup>With the following intuition. Consider crisp sets,  $A, B$ , subsets of some universe  $\Omega$ . Then  $B \subset A$  if and only if  $A \cup B' = \Omega$ . Hence we need to employ OR and NOT operators, and we use the min-max ones (cf. Smithson, 1987).

- 1 • if  $\mu_{fWTP}$  and  $\mu_{fWTA}$  are continuous and also  $i^{**}$  maximizes MNB, then  $fNB_{i^*}$   
2 and  $fNB_{i^{**}}$  weakly include each other to the same degree.

3 Maximizing MNB can be seen (not pursued formally, for brevity) as applying  
4 the Orlovsky-score (1978), i.e. maximizing the degree to which a given alternative  
5 is not dominated by any other. There is still additional intuition behind MNB,  
6 when thinking in terms of example in Figure 3 and options  $A$ – $D$ , with increasing  
7  $e$ . For options  $A$  and  $B$  the decision maker is convinced to a degree of  $>0.5$  it is  
8 worth to switch to a more effective option, while option  $D$ —convinced it is worth  
9 to switch to a less effective one. Only for  $C$  no such conviction prevails.

10 Maximizing MNB can be easiest done by estimating the upper bound of the  
11 0.5-cut for  $fWTP$  and  $fWTA$  and using these (crisp) values to calculate the, then  
12 crisp, NB. For each  $i$  we calculate  $NB_i = e_i \times \sup A_{fWTP}(0.5) - c_i$  (if  $e_i \geq 0$ ). In  
13 section 4 I propose three methods how to evaluate these 0.5-cuts for  $fWTP/fWTA$ .

14 Finally notice that other percentiles ( $\alpha$ -cuts of  $fNB$ ) could be used, but again  
15 requiring an interval interpretation. Taking  $\alpha > 0.5$  would effectively mean taking  
16 lower WTP but greater WTA values, i.e. the fanning out in CE-plane (Obenchain,  
17 2008, and J&K). I.e. if increasing WTP is to represent being more permissive in  
18 switching from *status quo* (or using a lower percentile in the present framework)  
19 in the I quadrant, then we need to accompany it with lowering WTA.

## 20 4 Calculating the 0.5-cut for fuzzy WTP & WTA

21 For brevity, call the 0.5-cut for  $fWTP/fWTA$  the *indecisiveness point (IP)*. IPs var-  
22 ied between respondents (horizontal bars scattered along the abscissa in Figure 2),  
23 and obtaining a single, population-level IP requires some aggregation, accounting  
24 for the randomness of the sample. Below I suggest three methods, using different  
25 approaches to statistical inference: hypothesis testing, Bayesian modelling, and  
26 frequentist estimation. The advantages and disadvantages are discussed, however,  
27 no clear winner is pointed. The last two methods require data transformation, de-  
28 scribed in subsection 4.2. The  $\lambda$ s denote the values used in the questionnaire and  
29 are presented in 000s PLN/QALY.

### 30 4.1 Hypothesis testing

31 The assumptions are presented for WTP, and are analogous for WTA. 1) For each  
32  $\lambda \in \mathbb{R}_+$  there is an (unknown) average conviction in the population,  $\mu_{WTP}(\lambda)$ . 2)  
33 Our estimand is IP such that  $\mu_{WTP}(IP) = 0.5$  (no uniqueness has to be assumed).  
34 3) Assume that the values of  $\mu_{WTP,i}(IP)$  for every individual,  $i$ , are drawn for a  
35 common, symmetric distribution, and so are the responses in the Likert scale.

1 For each  $\lambda$  we can test  $H_0 : IP = \lambda$ , testing the symmetry of the distribution  
2 of answers. I used the test suggested by Dykstra et al (1995) (with  $H_2$  as the  
3 alternative hypothesis, according to their notation). Mann-Whitney test could also  
4 be used (comparing the actual responses to vector of 3s); with no impact on the  
5 conclusions in the present data. Dykstra et al (1995) test seems to be using more  
6 information from the data (Mann-Whitney not differentiating between 1 and 2 or 4  
7 and 5 options), but the comparison of these (and other) tests should be performed  
8 when data have been collected.

9 For WTP we do not reject  $H_0$  for  $\lambda = 125$  ( $p^* = 0.0612$ ) and  $\lambda = 150$  ( $p^* =$   
10  $0.6313$ ), while e.g. for  $\lambda = 100$  or  $\lambda = 175$  we get  $p^* = 0.0001$  and  $p^* = 0.0028$ ,  
11 respectively. For WTA, we do not reject  $H_0$  for  $\lambda = 150$  ( $p^* = 0.1994$ ),  $\lambda = 175$   
12 ( $p^* = 0.2532$ ),  $\lambda = 200$  ( $p^* = 0.166$ ),  $\lambda = 250$  ( $p^* = 0.1308$ ), and  $\lambda = 300$  ( $p^* =$   
13  $0.0849$ ). The conclusions (which  $H_0$  are rejected) do not change if we double the  
14  $p^*$  values to account for one-sidedness of the alternative hypothesis. As we infer  
15 separately for each  $\lambda$ , there is no need to correct for multiple hypothesis testing.

## 16 4.2 Data transformation

17 Above I analysed each  $\lambda$  separately, but looked at the respondents' jointly. In  
18 two remaining approaches I proceed conversely: I consider each respondent in-  
19 dividually, looking at all the  $\lambda$ s for which the middle Likert option was chosen  
20 (interpreted as  $\mu(\lambda) = 0.5$ ) simultaneously. I call this range of  $\lambda$ s an *indecisive-*  
21 *ness range* (IR), and will use IR to estimate a single IP value.

22 Identifying IRs requires data transformation and assumptions, described below  
23 for WTP (analogous for WTA). Firstly, if the respondent did not use the middle  
24 option, I assume  $IR \neq \emptyset$  (simply no  $\lambda \in IR$  was used in the survey). I assume that  
25 option 3 would be used for  $\lambda$  equal to the average of the greatest  $\lambda$  with options 4  
26 or 5 selected and the lowest  $\lambda$  with 1 or 2.

27 Secondly, I assume IR's lower endpoint as the mean of the greatest  $\lambda$  with  
28 options 4 or 5 and the lowest  $\lambda$  with 3 (directly selected or inferred as above);  
29 analogously for the upper endpoint. Example 1: if the respondent selected option  
30 4 for  $\lambda = 100$ , option 3 for  $\lambda = 125$  and  $\lambda = 150$ , and option 2 for  $\lambda = 175$ , then  
31  $IR = [112.5; 162.5]$ . Example 2: if the respondent selected option 4 for  $\lambda = 100$   
32 and immediately switched to option 2 for  $\lambda = 125$ , then  $IR = [106.25; 118.75]$ .

33 The assumptions suffice to calculate IRs for WTP. In case of WTA, however,  
34 two respondents used only options 1 & 2, and one respondent only option 3, for  
35 all the  $\lambda$ s, thwarting the calculation of IR. I removed all three from the sample,  
36 based on two reasons. 1) These respondents do not conform to the *criteria trade-*  
37 *ability* axiom of J&K: they seem to, in principle, disagree that the decision maker  
38 should sometimes sacrifice effectiveness to make savings. The decision support  
39 methods developed in the present paper accept such trade-offs (and aim to express



1 them quantitatively), and should not be based on the opinions in such a funda-  
 2 mental disagreement with the foundations. 2) We aim here mostly to illustrate  
 3 the estimation methods, and not to come up with ultimate, ready-to-use estimates.  
 4 The latter would demand a larger sample and probably using more  $\lambda$ s in the ques-  
 5 tionnaire (or a possibility to freely report large values if the scale is insufficient).  
 6 Still, a further research is needed to consider how this, effectively infinite, WTA  
 7 should be accounted for quantitatively (how to combine finite and infinite WTA  
 8 in a mathematical framework).

9 Finally, I took the log of  $\lambda$ s (1 PLN/QALY added, to avoid  $\ln(0)$ ), for three rea-  
 10 sons. Firstly, the distribution of the middles of the IRs was highly skewed (skew-  
 11 ness coefficient 3.77 for WTP and 1.75 for WTA for non-log data, and 0.14 and  
 12 0.68, respectively, for logs), and statistical methods typically work on non-skew  
 13 data better. Secondly, the length of IR is positively correlated with the middle of  
 14 IR (for non-logs). Intuitively, the respondents thinking large allow larger tolerance  
 15 in absolute terms; plus  $\lambda$ s were more sparsely located for large values. It is more  
 16 convenient to model the respondents uncertainty in relative terms, not having to  
 17 directly model the relation between the IR's middle and length, and this is auto-  
 18 matically done with logs. Thirdly, with logs the results will not change whether  
 19 we use WTP/WTA expressed as a monetary value of a unit of health (PLN/QALY)  
 20 or a health equivalent of a monetary unit (QALY/PLN); not the case with original  
 21 data (arithmetic and geometric means not equivalent).

### 22 **4.3 Hierarchical Bayesian modelling**

23 In this and next subsection I use the following notation. For each of  $n$  respondents,  
 24 indexed by  $i \in \{1, \dots, n\}$ , we observe  $l_i$  ( $u_i$ ) denoting the lower (upper) endpoint  
 25 of IR (logs). Let  $m_i$  denote the middle of IR (i.e.  $m_i = (l_i + u_i)/2$ ). In short, in the  
 26 Bayesian approach we assume some statistical model how the observables are  
 27 generated from parameters of interest (with some prior distributions). We then  
 28 update the prior distributions based on actually observed values (for a description  
 29 and examples see, e.g. Ntzoufras, 2009).

30 Specifically, assume the following. Each respondent has a single, *true*, log of  
 31 indecisiveness point, denoted by  $\eta_i$ , drawn from a common distribution  $N(\eta, \xi^2)$ ;  
 32 taking the logs, conveniently, allows using a normal distribution, as the non-log IR  
 33 are bounded by zero from below. Then  $\eta$  is the main parameter of interest, allow-  
 34 ing to calculate  $\exp(\eta)$ . The respondent does not precisely perceive  $\eta_i$ , but rather  
 35 the bounds,  $l_i$  and  $u_i$ , generated as  $l_i = \eta_i - \Delta'_i$  and  $u_i = \eta_i + \Delta''_i$ , where  $\Delta'_i$  and  
 36  $\Delta''_i$  are independent random variables drawn from a single (for every respondent),  
 37 exponential distribution,  $\text{Exp}(\kappa)$ . The above statistical model defines the distri-  
 38 bution of observables  $(l_i, u_i)$  based on parameters  $(\eta, \xi, \kappa)$ . With a larger sample  
 39 we might consider assuming idiosyncratic  $\kappa$ s generated from some distribution.

1 The independence of  $\Delta$ s reflects the unpredictability of misjudging one's IP.  
 2 Using the exponential distribution has two nice consequences. Firstly, this dis-  
 3 tribution is memoryless—here implying: knowing that one's IP is misjudged up-  
 4 wards by at least some amount does not change the distribution of by how much  
 5 *more* this IP is misjudged. This reflects the lack of regularity in perceiving one's  
 6 IP. Secondly, the resulting distribution of  $\Delta'_i/(\Delta'_i+\Delta'_i)$  is uniform, and so the relative  
 7 location of the true value is non-informative, a reasonably conservative approach,  
 8 again, reflect the difficulty with positioning one's IP.

9 I used non-informative priors (normal for  $\mu$ , gamma for  $\sigma^2$  and  $\lambda$ ) and es-  
 10 timated the model with MCMC in JAGS/R (10,000 burn-in iterations, 50,000 of  
 11 actual iterations, thinning 5). The mean of the posterior was taken as the estimate,  
 12 and percentiles 2.5% and 97.5% as boundaries of the 95% credible interval (CI).

13 For WTP the estimate of  $\exp(\eta)$  equals 145.68, 95%CI = (106.99; 197.95),  
 14 while for WTA we get 162.29 and (115.78; 228.15), respectively. For the sake of  
 15 section 5: the posterior distribution of  $\eta$  was unimodal, symmetric, and leptokur-  
 16 tic (excess kurtosis equal to 0.53). The Shapiro test rejects normality ( $p^* < 0.001$ ).

#### 17 **4.4 A meta-analytic approach with bootstrap**

18 Here I employ the approach commonly used, e.g. to average the results of multi-  
 19 ple randomized clinical trials (see, e.g. Whitehead, 2002). I assume the random  
 20 effects model: respondents differ in terms of their *true* IP, denoted by  $\eta_i$ , drawn  
 21 from a  $N(\eta, \xi^2)$  (I use the same symbols as in the previous subsection to make it  
 22 easier, as some intuitions are identical). In the frequentist approach here,  $\eta$  is the  
 23 true, unknown parameter of interest (with no probability distribution).

24 I assume the precision for each  $i$  is given by the length of IR (I take the ob-  
 25 served length to be the actual precision, not accounting for the error of precision  
 26 estimation). I assume that observed IR ( $[l_i, u_i]$ ) is uniformly distributed in the real  
 27 axis, subject to  $\eta_i \in [l_i, u_i]$ . Then  $m_i = (l_i+u_i)/2$  is uniformly distributed around  $\eta_i$   
 28 with variance  $(u_i-l_i)^2/12$ , and  $m_i$  is an unbiased point estimate of  $\eta_i$  for every  $i$ . I  
 29 use the inverse-variance weighted average to calculate the point estimate  $\hat{\eta}$ , ac-  
 30 counting for random effects, using standard formulae (see, e.g. Whitehead, 2002).

31 The formulae are typically used for normal distributions, but are correct for  
 32 a uniform distribution and allow to estimate the random-effect variance from the  
 33 observables. Still, the distribution of estimated  $\hat{\eta}$  is not normal. For this reason I  
 34 assess the 95%confidence interval (CI, with a slight abuse of notation) for  $\eta$  via  
 35 bootstrapping (cf. Efron and Tibshirani, 1993): i) re-sample the set of respondents  
 36 (to account for sampling error), ii) for each re-sampled respondent generate a new  
 37  $m_i^*$  from a uniform distribution  $[l_i, u_i]$ , iii) keep the length of IR, iv) calculate the  $\hat{\eta}^*$   
 38 in this bootstrap sample (inverse-variance, random effects), v) repeat for 10,000  
 39 bootstrap samples and take percentiles 2.5% and 97.5% to define the 95%CI.

1 For WTP the  $\exp(\hat{\eta}) = 153.57$ ,  $95\%CI = (121.19; 202.89)$ . For WTA, respec-  
2 tively, 163.29 and  $(120.94; 225.13)$ . No problem of a bias is present, as the mean  
3 of bootstrap results yields 154.26 and 163.03 for WTP and WTA, respectively  
4 (very close to the meta-analysis results). Notice, that assuming the normal dis-  
5 tribution of the standard error in the meta-analysis would yield more narrow (and  
6 probably wrong)  $95\%CI$ :  $(130.57; 180.61)$  and  $(135.03; 197.47)$ , respectively. For  
7 the sake of section 5: the bootstrap distribution of  $\hat{\eta}$  was unimodal, slightly posi-  
8 tively skewed (0.33), and leptokurtic (excess kurtosis equal to 0.43). The Shapiro  
9 test rejects normality ( $p^* < 0.001$ ).

## 10 4.5 Comparison of approaches

11 Table 1 summarizes the—reassuringly consistent—results. The IP for WTP/WTA  
12 is greater than the official threshold in Poland (125,955 PLN/QALY as of 1st Nov,  
13 2015, and 111,381 PLN/QALY in the time the survey was run). There is no reason  
14 to believe that IP for WTA is greater than for WTP (increasing the sample size  
15 might change that conclusion), but all three methods suggest that there is more  
16 uncertainty for WTA.

*Table 1: Estimation results for the indecisiveness point (in 000s PLN/QALY) along with 95% confidence or credible (depending on context) interval (95%CI).*

Method	Willingness-to-pay	Willingness-to-accept
hypothesis testing	not rejected for 125, 150	not rejected for 150–300
Bayesian modelling (95%CI)	145.7, (107.0; 197.9)	162.3, (115.8; 228.1)
meta-analysis (95%CI)	153.6, (121.2; 202.9)	163.3, (120.9; 225.1)

17 The statistical testing requires fewest assumptions (e.g. no specific distribution  
18 assumed) and its results do not require (or change with) any transformation of  
19  $\lambda$ s; the other two methods would yield larger values if applied to original  $\lambda$ s,  
20 but still, taking the logs was motivated. Hypothesis testing works on complete  
21 data, while other methods have problems with respondents not crossing the middle  
22 Likert option. Also, the extremely undecided respondents (selecting the middle  
23 option option), when added to the sample, would change the results of the last two  
24 methods, while are effectively ignored by hypothesis testing.

25 The last two methods require the middle option, so as to account for possi-  
26 bly wide IR (otherwise we would underestimate the intra-respondent uncertainty  
27 regarding the location of IP). The middle option cannot, however, be too inclu-  
28 sive (e.g. *neither entirely agree, nor entirely disagree* in a 3-level Likert) for the  
29 last two methods, as that would change the estimand. The hypothesis testing ap-  
30 proach can be used irrespectively whether or not the middle level is used, how

1 it is worded (e.g. *neither/nor* or *I don't know*, as long as it is symmetric, i.e. not  
2 leaning towards agreeing or disagreeing), and whether it is framed to be more at-  
3 tractive for respondents. Still, making the middle answer too attractive reduces  
4 the power of the test, as fewer observations constitute the actual sample for the  
5 Dykstra et al (1995) test. Matell and Jacoby (1972) showed that using more (odd  
6 number of) levels decreases the frequency of selecting the middle option—hence.  
7 All the methods might profit from using a greater than five, odd number of lev-  
8 els. Even though the levels might then lose natural interpretation (section 2.2), if  
9 respondents symmetrically behave on two sides of the middle option that would  
10 increase the precision of IP estimation.

11 None of the methods interpreted the Likert scale as an interval one. In the  
12 hypothesis testing we do, however, assume that options 1 & 2 are symmetrical  
13 counterparts of 5 & 4. This does not seem to be a strong assumption, as the  
14 wording is symmetrical. Hypothesis testing can only be used to assess IP, while  
15 the other two methods could be adapted to estimate the range of  $\lambda$ s for which,  
16 e.g.  $\mu(\cdot) = 0.75$  (interpreting Likert scale in that way).

17 The usefulness of hypothesis testing depends most heavily on the number of  
18  $\lambda$ s used in the questionnaire: for WTA we did not reject  $H_0$  for  $\lambda = 300$ , and reject  
19 it for 400, with a wide range of untested values. Using more  $\lambda$ s would be tire-  
20 some for the respondents, and increasing the density, e.g. around 150 PLN, could  
21 bias the respondents towards locating IP in this region, suggesting that something  
22 should be happening there (a form of a central tendency bias). This is the biggest  
23 downside to the hypothesis testing approach.

24 The outcome of hypothesis testing may be disappointing for some. Not re-  
25 jecting  $H_0$  does not denote accepting it in statistical parlance. We also have to  
26 treat all the non-rejected  $\lambda$ s in the same way—there is no telling which are more  
27 likely to actually represent IP (one might try to use  $p^*$  for this purpose, adopting a  
28 Fisherian rather than Neyman-Pearson approach, a discussion beyond the scope of  
29 this paper). Bayesian approach conveniently produces a posteriori distributions,  
30 to be easily used in sensitivity analysis (cf. section 5). It could also account for  
31 covariates and explain part of the heterogeneity between the respondents.

32 Interpreting the Likert scale in a stronger way would allow to define other  
33 approaches. One could assume some (most likely, S-shape) parametric function  
34 how the 1–5 Likert responses change with  $\lambda$ , and estimate the parameters based  
35 on all observed responses (a form of the Rasch model could be used). This would  
36 be required to be then able to calculate  $\beta$ s and ANBs as defined in section 3.2.

37 Finally, we might try to explicitly allow for  $\mu_{fWTP}(x) = 0.5$  ( $\mu_{fWTA}(x) = 0.5$ )  
38 for a range of  $x$ . Looking at the definition of MfNB we now need to find the  
39 largest such  $x$  for fWTP and the smallest for fWTA. We don't need to change  
40 anything in the hypothesis testing approach, and then, conveniently, we infer that  
41 this value amounts to 150 (000s PLN/QALY) for both WTP and WTA. For the re-

1 maining two approaches we would be interested in the respective ends of the IR,  
2  $u_i$  or  $l_i$ . It is difficult to naturally quantify the individual-level error of  $u_i$  ( $l_i$ ),  
3 and so the most natural approach is to average the observed values, and account  
4 for the sampling error with, e.g. bootstrap over the respondents. We then obtain  
5 (in 000s PLN/QALY) 165.96 with 95%CI = (131.21;219.49) for WTP, and 140.53  
6 with 95%CI = (103.84;197.09) for WTA. However numerically greater WTP is  
7 WTA, the difference is not statistically significant (even a 90%CI for difference  
8 contains 0). Hence, whatever approach we take, there seems to be no rationale to  
9 systematically differentiate WTP from WTA.

## 10 5 Uncertainty & sensitivity analysis

11 The fuzzy framework, apart from suggesting a new decision making rule, allows a  
12 form of SA not referring to stochastic uncertainty. In spite of choosing with MNB,  
13 parameters  $\beta_i$  (subsection 3.2.1) illustrate whether the choice is best *per se* (large  
14  $\beta$ ) or is a compromise between too costly and too ineffective alternatives (low  $\beta$ ).  
15 For example ( $D_i$  ranked by  $e_i$ ) whether  $\beta_1 = 0.2$ ,  $\beta_2 = 0.7$ ,  $\beta_3 = 0.1$  or  $\beta_1 = 0.45$ ,  
16  $\beta_2 = 0.1$ ,  $\beta_3 = 0.45$  using MNB selects  $D_2$ , but the story behind differs.

17 The three decision making rules start to agree, when fuzziness is reduced, as  
18 the following proposition states (presented for the right half of CE-plane but holds  
19 for the whole plane).

20 **Proposition 2.** Consider a sequence of fuzzy numbers,  $fWTP^{(j)}$ , converging to a  
21 crisp number WTP, i.e.  $\sup A_{fWTP^{(j)}}(0) \rightarrow WTP$  and  $\sup A_{fWTP^{(j)}}(1) \rightarrow WTP$ , limits  
22 when  $(j) \rightarrow +\infty$ . Take  $n$  technologies,  $(e_i, c_i) \in \mathbb{R}_+ \times \mathbb{R}$ , such that there is a single  
23 technology,  $i^*$ , maximizes  $NB_i = WTP \times e_i - c_i$ . Then  $\forall \varepsilon > 0 \exists M \in \mathbb{N}$ , such that  
24  $\forall j > M$  if we calculate  $MNB_i$ ,  $ANB_i$ , and  $\beta_i$  for  $fWTP^{(j)}$ :

- 25 1.  $i^*$  maximizes  $MNB_i$ , as the only technology,
- 26 2.  $i^*$  maximizes  $ANB_i$ , as the only technology, and  $|ANB_{i^*} - MNB_{i^*}| < \varepsilon$ ,
- 27 3.  $i^*$  maximizes  $\beta_i$ , as the only technology, and  $\beta_{i^*} > 1 - \varepsilon$ .

28 In HTA there is stochastic uncertainty, and SA is used to illustrate its im-  
29 pact. The main source is that  $(e_i, c_i)$  are estimated based on clinical trials, meta-  
30 analyses, or pharmacoeconomic models (cf. Briggs et al, 2012). The current  
31 framework, adds two elements. Firstly, the IPs are also estimated, and so us-  
32 ing MNB removes fuzziness in the end, introducing more stochastic uncertainty.  
33 Secondly, the selection of baseline technology (with respect to which  $(e_i, c_i)$  are  
34 calculated) impacts the results by moving the points between halves of the CE-  
35 plane, important when  $fWTP$  and  $fWTA$  differ. Often a mix of technologies will

1 be used. Then we can calculate  $(e_i, c_i)$  with different baseline technologies in this  
 2 mix, with proper market shares taken as weights/probabilities. This uncertainty  
 3 can be simply joint with  $(e, c)$ -estimation uncertainty (but will be left out, for  
 4 clarity, in the examples that follow).

5 We will (as often done) approach uncertainty in a Bayesian way: assume a dis-  
 6 tribution of model parameters, from which to draw  $(e_i, c_i)$  and IPs to use in Monte  
 7 Carlo analysis. The IPs will be drawn independently from  $(e_i, c_i)$ . Whether IPs for  
 8 fWTP and fWTA should be independent can be disputed. On one extreme, they  
 9 were elicited and estimated separately, suggesting independence. On the other  
 10 extreme, there is no statistical reason (in our case) to reject their equality, hence  
 11 they should be assumed equal. In between, the individuals assessing WTP to be  
 12 large, tend to assess WTA to be large, and so sample randomness still suggests  
 13 some positive correlation between the two. If IPs are not equalized by definition,  
 14 then randomizing *status quo* impacts the results.

15 To present a wider context, how to change WTA with WTP depends on the  
 16 goal. We want them to change in the same direction, to represent uncertainty on  
 17 the actual value and we want to introduce a positive correlation in estimation error.  
 18 Severens et al (2005) suggested to, when performing SA with cost-effectiveness  
 19 acceptability curves (CEACs), keep the  $WTA/WTP$  constant, and change WTP (hence,  
 20 changing WTA in the same direction). On the other hand, we want to change WTP  
 21 and WTA in the opposite directions, to represent a changing decision making rule.  
 22 To be more permissive in our decisions (lowering the  $\alpha$  in  $\alpha$ -cut for fNB), then  
 23 we should *fan out*: use higher WTP and lower WTA.

24 Back to SA, as a result we get, for each  $i$ , the empirical distribution of  $MNB_i$ .  
 25 (Notice, that had we estimated the complete  $\mu_{fWTP}, \mu_{fWTA}$ , we could also consider  
 26 the  $\beta$ s and ANBs.) The next proposition shows the properties of this random  
 27 variable when uncertainty is reduced, to make sure that accounting for uncertainty  
 28 is a natural extension, i.e. MNB behaves in a predictable way.

29 **Proposition 3.** *Take a sequence of random variables  $(e^{(j)}, c^{(j)}, IP_{fWTP}^{(j)}, IP_{fWTA}^{(j)})$ ,*  
 30 *with  $(j)$  numbering elements. Assume the sequence converges in probability to a*  
 31 *degenerate distribution located in  $(e, c, IP_{fWTP}, IP_{fWTA})$ . For each  $(j)$  define a ran-*  
 32 *dom variable  $MNB^{(j)}$  as in eq. 4. Then the sequence  $MNB^{(j)}$  converges in proba-*  
 33 *bility to a degenerate random variable: MNB calculated for  $(e, c, IP_{WTP}, IP_{WTA})$ .*

34 The insights gained by analysing the distributions of MNBs is illustrated via  
 35 two examples. Start with one presented in Figure 3. Assume that uncertainty re-  
 36 garding each  $(e_i, c_i)$  is given by a bivariate normal distribution, around the means  
 37 in the Figure, with SDs equal to 0.2 (and no correlation). Assume  $IP_{fWTP} =$   
 38  $\exp(\eta)$  with  $\eta \sim N(\ln(1.5), 0.25)$ . Some points may fall in quadrants II & III,  
 39 so we consider fWTA, distributed as fWTP. We disregard  $E$  (used only to show  
 40 the Chernoff property violation).

1 As a reminder: point estimates led to choosing  $C$ , with  $MNB_C = 3 \times 1.5 -$   
2  $2 = 2.5$ , while, e.g.  $MNB_D = 2.45$ ;  $\beta_C = 0.3$  was not the greatest ( $\beta_D = 0.45$ ),  
3 showing that  $C$  was a compromise. Now, the average MNB (for 10,000 Monte  
4 Carlo draws) equals: 0.41, 2.41, 1.30, 2.35, 2.64, and 2.63, for  $X$ ,  $Y$ ,  $A$ ,  $B$ ,  $C$ ,  
5 and  $D$ , respectively. The asymmetry of distributions in Monte Carlo makes  $C$   
6 almost equal to  $D$  now, the log-normal distribution results in average of  $\exp(\eta)$   
7 being greater than  $\exp(1.5)$ , and so the average MNBs may differ from the ones  
8 calculated in baseline analysis. Larger uncertainty might result in average MNB  
9 larger for  $D$  than  $C$ , introducing a discordance between the baseline and sensitivity  
10 analysis (possible also for standard SA in HTA, when skewness is present).

11 Just like using CEACs, a popular tool in SA (for more information see, e.g.  
12 van Hout et al, 1994; Fenwick et al, 2004), we can calculate the probability of  
13 each  $i$  maximizing MNB. Technically, we would obtain similar result averaging  
14 the CEAC values over horizontal axis with weights taken as probability distribu-  
15 tion for WTP. The difference would stem from lack of possibility to differentiate  
16 between WTP and WTA by regular CEACs (see Severens et al, 2005; Araki and  
17 Kamae, 2015, for some ideas). The probability of maximizing MNB amounts to  
18 0%, 16.2%, 0.7%, 19.5%, 29.2%, and 34.5%, for  $X$ ,  $Y$ ,  $A$ ,  $B$ ,  $C$ , and  $D$ , respec-  
19 tively. Hence, just like with regular CEACs, the probability driven results need not  
20 agree with expectation driven Fenwick et al (2001). The approaches would agree  
21 in the limit when uncertainty is being reduced (Proposition 3). Interestingly, this  
22 discordance would not occur here for regular CEACs due to the normality of  $(e, c)$   
23 distribution and lack of correlation (Jakubczyk and Kamiński, 2010; Sadatsafavi  
24 et al, 2008), and is only introduced by additional uncertainty of WTP and WTA.

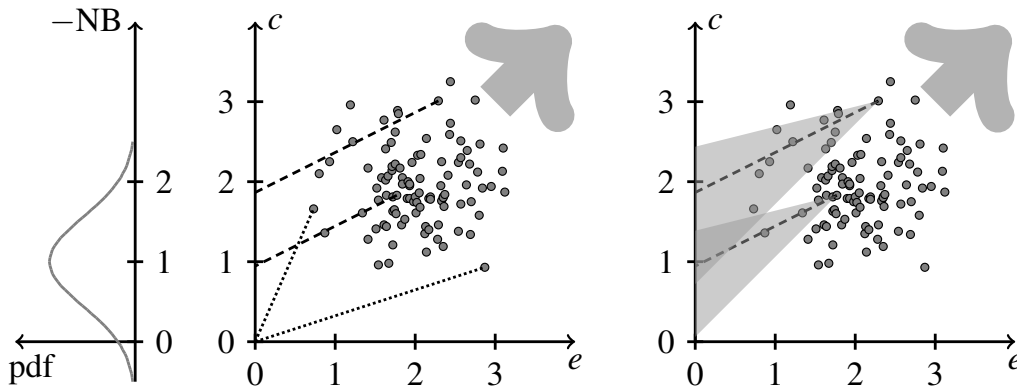
25 So as to present the properties of SA in the current framework, consider a  
26 simpler example. Take the means of  $(e_i, c_i)$  to be  $D_1 = (-1, -1)$ ,  $D_2 = (0, 0)$  (ex-  
27 plicitly modelling the null option), and  $D_3 = (1, 1)$ . Assume  $(e_i, c_i)$  are normally  
28 and independently distributed with all  $SD = 0.1$ . Take  $IP_{fWTP}$ ,  $IP_{fWTA}$  to be iden-  
29 tically, independently distributed as exp of the underlying distribution  $N(0, 0.2)$ .

30 The skewness of the  $fWTP/fWTA$  results in an asymmetric treatment to effect-  
31 improving and -reducing technologies. Mean MNBs are almost equal: -0.02, 0,  
32 and 0.02 (for  $D_1-D_3$ ), but the probabilities of maximizing MNB amount to, re-  
33 spectively, 33.7%, 29.4%, and 36.9%. There are two reasons. Firstly, the distri-  
34 bution of MNB spreads more the farther away a given technology,  $D_i = (e_i, c_i)$ , is  
35 from the  $y$ -axis, as uncertainty on  $fWTP/fWTA$  results in a bigger spread when  $e_i$   
36 is recalculated to monetary values. Secondly, due to the log-normal distribution  
37 of IPs, the the distribution of MNBs is left-skew for  $D_1$  and right skew for  $D_3$ , and  
38 the long right tail helps  $D_3$  to maximize the probability.

39 It is most informative to calculate some low percentiles of MNB distributions,  
40 to see how risky the available alternatives are. For example, the 5% percentile  
41 is equal to -0.47, -0.23, and -0.35 (for  $D_1-D_3$ ), and so the risk averse decision

1 maker should favour the null option. Analysing the percentiles of MNB has a  
 2 nice property that moving away from the origin in CE-plane (to be more precise:  
 3 increasing the absolute value of  $e$ ) increases the risk. That seems intuitive, as  
 4 implies that larger trade-offs are being made, which should feel risky for a decision  
 5 maker uncertain of own WTP/WTA. That is a new property, absent in standard  
 6 CEAC analysis (and criticized, e.g. by Koerkamp et al, 2007).

7 This is illustrated in Figure 4. When WTP is given as parameter, then mov-  
 8 ing the cloud of points (representing uncertainty) away from the y-axis does not  
 9 spread the distribution of NB (density function, pdf, illustrated in the left panel), as  
 10 calculating NB can be visualised as projecting the scattered points on the y-axis  
 11 along lines with slope representing WTP. When WTP is given with uncertainty  
 12 (right panel), then the projection is done along lines fanning out (represented by  
 13 gray areas), and moving the cloud of points results in wider spread.



*Figure 4: Increasing the absolute value of effect does not change the total uncertainty of NB when WTP is given (in central panel, left panel represents the density function of NB) and increases it when WTP is estimated with uncertainty (right panel).*

14 As mentioned above, the probability of maximizing MNB could be (almost)  
 15 read off the standard CEAC, by averaging over the range of WTPs. That is not  
 16 the case for expected value of perfect information (EVPI). An example below  
 17 illustrates that the uncertainty in IPs for WTP and WTA results in qualitatively  
 18 new phenomena, that cannot be seen in standard graphs used to illustrate EVPI.  
 19 Of course the very amount of uncertainty is larger, and so EVPI increases, but the  
 20 difference is also a qualitative one. Consider comparing  $(e, c) = (k, k)$  with SD  
 21 equal to 1 (for  $e$  and  $c$ ) for  $k = 1, 2, 5, 10, 20$  versus  $(0, 0)$ , i.e. consider a set of five  
 22 decision problems. Larger  $k$  represents larger shift outwards. In Figure 5 the EVPI  
 23 are presented for various WTP. As  $k$  increases (represented by a darker shade),  
 24 EVPI seems to decrease for all but one value of WTP, giving the impression that  
 25 overall there is less uncertainty in the problem. When we assume WTP is only



1 given as a distribution (here, a uniform  $[1/2, 3/2]$ ), then the resulting EVPI (now, a  
 2 single number) increases with  $k$ : 0.59, 0.61, 0.8, 1.36, and 2.55.

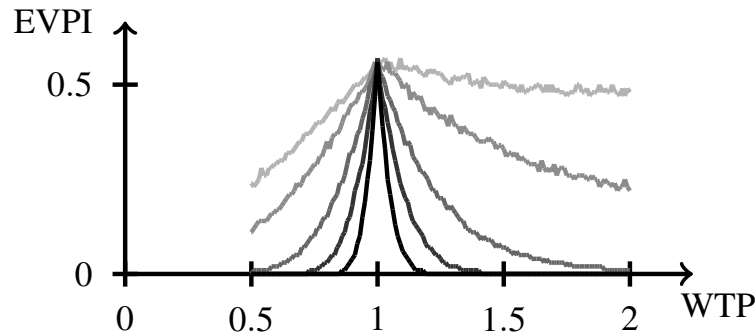


Figure 5: EVPI for a comparison of  $(e, c)$  distributed around  $(k, k)$  vs  $(0, 0)$  when  $k = 1, 2, 5, 10, 20$  (darker shade, larger  $k$ ).

## 3 6 Concluding remarks

4 In the paper I tried to comprehensively show how to make the fuzzy approach to  
 5 modelling WTP/WTA operational, i.e. how to build a complete decision making  
 6 process, along with methods how to estimate model's parameters and conduct SA.  
 7 The framework works for multiple alternatives, effect improving or reducing, and  
 8 can be combined with other types of stochastic uncertainty. Apart from (crisp)  
 9 choices the model yields via SA additional information on decision robustness  
 10 and can be used along regular, crisp CEA. Importantly, parts of the present paper  
 11 (e.g. estimation techniques) can be applied on their own.

12 Fuzzy approach can be discredited, as introducing too much subjectivity. In  
 13 the defence, the subjective notions are commonly used in CEA, e.g. assigning  
 14 utilities requires patients subjectively determining their quality of life via ques-  
 15 tionnaires. If carefully elicited and not wilfully biased, the subjective notions can  
 16 be used to inform better decisions. Additionally, the core of the methodology in  
 17 the paper uses the middle Likert answer, the least ambiguously defined.

18 What is achieved by introducing fuzziness, if it's dropped in the final choice?  
 19 At least four things. Face validity for one: if WTP/WTA is perceived fuzzily, then  
 20 the arguments are rather needed *not to use it*. Using fuzzy WTP/WTA, as long  
 21 as possible, matches the actual process of thinking better and the primitives of  
 22 the model are more strongly ontologically grounded. Secondly, building on con-  
 23 cepts independently developed in fuzzy-set theory (e.g. weak inclusion) allowed  
 24 to formally motivate useful simplifications—basing the choice on (crisp) IPs for

1 WTP/WTA. Even though using IPs seems as disregarding fuzziness, had it not  
2 been for the fuzzy approach, we wouldn't have been able to even define the IP.

3 Thirdly, estimating the membership function for fuzzy WTP/WTA necessarily  
4 involves statistical uncertainty. Handling this estimation error is only possible in  
5 the fuzzy model within which it originates. Then the uncertainty (combined with  
6 other forms of uncertainty) can be used in SA to inform about the robustness of  
7 a given, crisp decision. The imprecise knowledge of WTP/WTA, when addressed  
8 in formal statistical inference fashion, yields new, intuitively-appealing, insight  
9 in SA: considering larger cost-effect trade-offs results in more uncertainty in the  
10 problem (not accounted for by standard CEAC or EVPI analysis).

11 Fourthly, the fuzzy approach allows to redefine the WTP/WTA disparity, and  
12 the proposed estimation methods allow to grasp it quantitatively. With the present  
13 data WTP-WTA disparity, when related to IPs, is not confirmed. Importantly, basing  
14 the decision on IP followed from different criteria, the eradication of disparity  
15 came as a convenient by-product. Still, if people base actual decisions on values  
16 closer to freely reported WTP/WTA (hence values with higher than 0.5 conviction,  
17 see section 2.2), then the elicited WTP and WTA will differ. This mechanism  
18 somewhat resembles the one introduced by Zhao and Kling (2001), who modelled  
19 the value of the good as given with stochastic uncertainty. The decision maker  
20 then plays it safe: decides to wait and collect information unless the price to  
21 pay is sufficiently low (to discourage from incurring the cost of waiting) or the  
22 price to accept is sufficiently high. In the fuzzy context: not certain about their  
23 perceptions, the respondents play safe and report values of higher conviction. This  
24 'playing safe' was also observed by Dubourg et al (1994): when having to select a  
25 single value from a range of possible WTPs/WTAs, the respondent selected point  
26 in the lower region for WTP (that is not confirmed by J&K's data, though) and  
27 higher region for WTA. Dubourg et al (1994) also observed that often the very  
28 regions for WTA were located higher than and did not overlap with regions for  
29 WTP, which fuzzy model explains as reporting  $\alpha$ -cuts for large  $\alpha$ s.

30 The paper can help to design questionnaires eliciting WTP/WTA. Likert scales  
31 seem more credible than and different from freely reported WTP/WTA values  
32 (observed by J&K). The neutral option must be used, so as to employ Bayesian  
33 or frequentist approach to estimate IP. More than 5 levels can be considered to  
34 improve the precision of hypothesis testing. Increasing the number of values ( $\lambda$ s)  
35 in the questionnaire improves the precision but tires the respondent. Perhaps using  
36 different sets could be considered to reduce the impact of gaps, but that prevents  
37 hypothesis testing, unless sample size is large. Finally, the respondents may have  
38 tried to answer symmetrically for WTA and WTP (even though asked not to go  
39 back to previous questions). It might be a good idea, then, to use two sets of  
40 questionnaires, starting from either WTP or WTA, to compare the results.

41 Several ideas for further research originate. Firstly, fuzzy *sets*, not *numbers*,

1 were used to describe fNB ( $\alpha$ -cuts were left-unbounded). Less technically, the  
2 interpretation of, e.g.  $\mu_{\text{fNB}}(5) = 1$  was: ‘*I’m fully convinced that (using a given*  
3 *HT) I gain at least 5 (in monetary terms)*’. A different approach would be to use  
4 fuzzy numbers and represent the opinions ‘*I’m ... convinced that I gain exactly*  
5 *...* ’. Secondly, the fuzzy set theory allows multiplication or addition of fuzzy  
6 values. Hence, the presented framework can also accommodate fuzzy measures of  
7 effectiveness, e.g. fuzzily perceived gains in quality of life. Thirdly, in the model I  
8 differentiated between the left and right halves of the CE-plane (i.e. between WTP  
9 and WTA). Why not divide between the upper and lower halves? Looking at the  
10 sign of effects, not costs, seems intuitive, yet lacks formal rationale. Perhaps the  
11 trade-off coefficient differ in all the quadrants, and these differences are simply  
12 overlooked in quadrants II & IV, as are overshadowed by dominance. Still, when  
13 performing SA, it is needed to also analyse quantitatively the part of distribution  
14 in all the quadrants to average out the results.

## 15 Acknowledgements

16 I would like to appreciate the Fulbright Senior Award, giving me the possibility to  
17 spend one semester at The University of Iowa and to conduct most of this research.  
18 Thanks also go to B. Kamiński for commenting on the earlier version of the paper.

## 19 Proofs

20 *Lemma 1.* It is intuitively straightforward (when looking at CE-plane), but let’s  
21 do the algebra. Take any  $\alpha \in (0, 1]$ . We focus on the right side of CE-plane, but  
22 all is analogous in the left side.

23 For  $\gamma > 0$ ,  $x \in A_{\text{fNB}(e,c)}(\alpha) \stackrel{\text{(i)}}{\Leftrightarrow} \mu(e, c+x) \geq \alpha \stackrel{\text{(ii)}}{\Leftrightarrow} \mu(\gamma e, \gamma c + \gamma x) \geq \alpha \stackrel{\text{(iii)}}{\Leftrightarrow} \gamma x \in$   
24  $A_{\text{fNB}(\gamma e, \lambda c)}$ , (i) and (iii) from the def. of  $\alpha$ -cut, and (ii) as  $\mu$  is constant on rays.

25 Now,  $A_{\text{fNB}(e_1, c_1+c)} \stackrel{\text{(i)}}{=} \{-c\} \oplus A_{\text{fNB}(e_1, c_1)}(\alpha) \stackrel{\text{(ii)}}{=} (-\infty, -c] \oplus A_{\text{fNB}(e_1, c_1)}(\alpha) \stackrel{\text{(iii)}}{=} A_{\text{fNB}(0,c)}(\alpha) \oplus A_{\text{fNB}(e_1, c_1)}(\alpha)$ , (i) from Def. 1, (ii) from  $\alpha$ -cuts being left-  
26 unbounded, and (iii) from the shape of  $\mu$  along the y-axis.

27 Consider adding  $(e, 0)$  to  $(e_1, c_1)$ , where  $e > 0 < e_1$  (otherwise back to the  
28 preceding paragraph). If  $\mu_{\text{fWTP}}(x/e) \geq \alpha$  and  $\mu_{\text{fWTP}}((c_1+x_1)/e_1) \geq \alpha$ , then also  
29  $\mu_{\text{fWTP}}((c_1+x_1+x)/(e_1+e)) \geq \alpha$  as  $(c_1+x_1+x)/(e_1+e) \leq \max\{x/e, (c_1+x_1)/e_1\}$ , and fWTP  
30 is non-increasing; hence,  $A_{\text{fNB}}(e, 0)(\alpha) \oplus A_{\text{fNB}}(e_1, c_1)(\alpha) \subset A_{\text{fNB}}(e_1 + e, c_1)(\alpha)$ .  
31 On the other hand, assume  $\mu_{\text{fWTP}}((c_1+x)/e_1+e) \geq \alpha$  and let  $c^* = e(c_1+x)/(e_1+e)$ ,  $c^{**} =$   
32  $e_1(c_1+x)/(e_1+e)$ . Clearly,  $\mu_{\text{fWTP}}(c^*/e) = \mu_{\text{fWTP}}(c^{**}/e_1) \geq \alpha$ , and  $c^* + (c^{**} - c_1) = x$ ;  
33 hence  $A_{\text{fNB}}(e_1 + e, c_1)(\alpha) \subset A_{\text{fNB}}(e, 0)(\alpha) \oplus A_{\text{fNB}}(e_1, c_1)(\alpha)$ .  
34 □

35

1 *Corollary 1.* Notice that  $(e_1, c_1) = (e_2, c_2) + (\Delta_c, \Delta_e)$ ,  $\Delta_e \geq 0$  and  $\Delta_c \leq 0$ , and so  
2  $0 \in A_{\text{fNB}(\Delta_e, \Delta_c)}(\alpha)$  for any  $\alpha \in (0, 1]$ ; then use  $\oplus$ . Strict version follows trivially. If  
3  $(e_1, c_1)$ ,  $(e_2, c_2)$  are separated by y-axis, then use  $(0, (c_1+c_2)/2)$  as an intermediary.  
4 □

5 *Corollary 2.* If  $\gamma \in \{0, 1\}$ , then we have the regular dominance. If  $A_{\text{fNB}(e_3, c_3)} \not\subset$   
6  $A_{\text{fNB}(e_1, c_1)} \cup A_{\text{fNB}(e_2, c_2)}$ , then  $\forall \gamma \in (0, 1)$ ,  $A_{\text{fNB}(e_3, c_3)} \not\subset \gamma \odot A_{\text{fNB}(e_1, c_1)} \oplus (1 - \gamma) \odot$   
7  $A_{\text{fNB}(e_2, c_2)}$ , and  $(e_3, c_3)$  can't be Pareto-dominated by the convex combination. □

8 *Proposition 1.* Proving the first part. Take any  $x \in (\text{MNB}_i, \text{MNB}_{i^*})$  (the inter-  
9 val is non-empty),  $\mu_{\text{fNB}(i)}(x) \leq 1/2 \leq \mu_{\text{fNB}(i^*)}(x)$ ; using the monotonicity of  $\mu_{\text{fNB}}$   
10 (for  $i^*, i$ ) yields the result. Proving the first bullet implication: take any  $x \in$   
11  $(\text{MNB}_i, \text{MNB}_{i^*})$  (again, exists),  $\mu_{\text{fNB}(i)}(x) < 1/2 < \mu_{\text{fNB}(i^*)}(x)$ , and use monotonic-  
12 ity again. Proving the last bullet. First consider  $e_{i^*} \neq 0 \neq e_{i^{**}}$ , and so  $\mu_{\text{fNB}}$  are con-  
13 tinuous for  $i^*, i^{**}$ . Then  $\mu_{\text{fNB}(i^*)}(x) = \mu_{\text{fNB}(i^{**})}(x) = 1/2$ , and the rest follows from  
14 monotonicity. Now consider  $e_{i^*} = 0 = e_{i^{**}}$ , then fNBs are equal, crisp numbers  
15 (with upper semi-continuous, step membership functions, jumping from 1 to 0),  
16 and so weakly include each other to the degree 1. Finally consider  $e_{i^*} \neq 0 = e_{i^{**}}$ .  
17  $\mu_{\text{fNB}(i^*)}$  is continuous and monotonic, and  $\text{fNB}_{i^{**}}$  is a crisp number. It easily fol-  
18 lows (considering  $x = \text{MNB}_{i^*}$ ) that  $\text{fNB}_{i^*}$  weakly includes  $\text{fNB}_{i^{**}}$  to the degree  $1/2$ .  
19 Approaching this  $x$  from right yields the weak inclusion to the same degree. □

20 *Proposition 2.* If for all  $i = 1, \dots, n$  we have  $e_i = 0$ , then technology  $i^*$  must have  
21 the lowest cost, and fWTP does not matter. Assume at least one  $e_i > 0$  (remember,  
22 we are in quadrants I & IV) and let  $\varepsilon = \text{NB}_{i^*} - \max_{j \neq i^*} \text{NB}_j$ . We may always find  
23  $m \in \mathbb{N}$ , such that for all  $(j) > m$  we have the suprema of  $\alpha$ -cuts not farther from  
24 fWTP than  $\varepsilon/2 \times \max(e_i)$ . Then any  $\alpha$ -cut of fNB for  $i^*$  is greater than respective  
25  $\alpha$ -cuts of other technologies. Increasing  $(j)$  we also get 0-cut arbitrarily close to  
26 (in Hausdorff metric) 1-cuts for fNB for  $i^*$ . These immediately yield 1–3. □

27 *Proposition 3.* Looking at the definitions of fNB (Def. 1),  $\alpha$ -cuts (eq. 1), and  $\tau$   
28 (eq. 4) we see that  $\tau = \sup A_{\text{fNB}(e, c)}(\alpha)$  is a continuous function of  $(e, c)$  in  $\mathbb{R}^2$ . By  
29 the continuous mapping theorem (cf. Billingsley, 1999) we get the result. □

## 30 **References**

31 Andersson H (2007) Willingness to pay for road safety and estimates of the risk of  
32 death: Evidence from a Swedish contingent valuation study. *Accident Analysis*  
33 *and Prevention* 39:853–865

- 1 Araki D, Kamae I (2015) The Augmented Representation of the Cost-  
2 effectiveness Acceptability Curve for Economic Evaluation of Health Technol-  
3 ogy. *Kobe Journal of Medical Sciences* 61(1):E9–E18
- 4 Bellavance F, Dionne G, Lebeau M (2009) The value of a statistical life: a meta-  
5 analysis with a mixed effects regression model. *Journal of Health Economics*  
6 28(2):444–464
- 7 Billingsley P (1999) *Convergence of Probability Measures*, 2nd edn. Wiley
- 8 Briggs A, Weinstein M, Fenwick E, Karnon J, Sculpher M, Paltiel A, on Behalf  
9 of the ISPOR-SMDM Modeling Good Research Practices Task Force (2012)  
10 Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM  
11 Modeling Good Research Practices Task Force-6. *Value in Health* 15:835–842
- 12 Doucouliagos C, TD S, Giles M (2012) Are estimates of the value of a statistical  
13 life exaggerated? *Journal of Health Economics* 31(1):197–206
- 14 Dubois D, Prade H (1980) *Fuzzy Sets and Systems: Theory and Applications*.  
15 New York: Academic Press
- 16 Dubourg W, Jones-Lee M, Loomes G (1994) Imprecise Preferences and the WTP-  
17 WTA Disparity. *Journal of Risk and Uncertainty* 9:115–133
- 18 Dykstra R, Kochar S, Robertson T (1995) Likelihood ratio tests for symmetry  
19 against one-sided alternatives. *Annals of the Institute of Statistical Mathematics*  
20 47(4):719–730
- 21 Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*, 1st edn. Chapman  
22 and Hall/CRC
- 23 Fenwick E, Claxton K, Sculpher M (2001) Representing uncertainty: the role of  
24 cost-effectiveness acceptability curves. *Health Economics* 10(8):779–787
- 25 Fenwick E, O’Brien B, Briggs A (2004) Cost-effectiveness acceptability curves—  
26 facts, fallacies and frequently asked questions. *Health Economics* 13:405–415
- 27 Garber A (2000) Advances in Cost-Effectiveness Analysis of Health Interven-  
28 tions. In: AJ Culyer JN (ed) *Handbook of Health Economics*. Volume 1A,  
29 North-Holland, pp 181–221
- 30 Grosse S (2008) Assessing cost-effectiveness in healthcare: history of the \$50,000  
31 per QALY threshold. *Expert Review of Pharmacoeconomics & Outcomes Re-  
32 search* 8(2):165–178

- 1 Horowitz J, McConell K (2002) A Review of WTP/WTA Studies. *Journal of Environmental Economics and Management* 44(3):426–447
- 2
- 3 van Hout B, Al M, Gordon G, Rutten F (1994) Costs, effects and C:E-ratios alongside a clinical trial. *Health Economics* 3:309–319
- 4
- 5 Hultkrantz L, Svensson M (2012) The value of a statistical life in Sweden: a review of the empirical literature. *Health Policy* 108(2–3):302–310
- 6
- 7 Jakubczyk M, Kamiński B (2015) Fuzzy approach to decision analysis with multiple criteria and uncertainty in health technology assessment. *Annals of Operations Research* DOI 10.1007/s10479-015-1910-9
- 8
- 9
- 10 Jakubczyk M (2016) Using a fuzzy approach in multi-criteria decision making with multiple alternatives in health care. *Multiple Criteria Decision Making* forthcoming:
- 11
- 12
- 13 Jakubczyk M, Kamiński B (2010) Cost-Effectiveness Acceptability Curves—Caveats Quantified. *Health Economics* 19:955–963
- 14
- 15 Johansson-Stenman O, Svedsäter H (2012) Self-image and valuation of moral goods: Stated versus actual willingness to pay. *Journal of Economic Behavior & Organization* 84:879–891
- 16
- 17
- 18 Koerkamp B, Hunink M, Stijnen T, Hammitt J, Kuntz K, Weinstein M (2007) Limitations of Acceptability Curves for Presenting Uncertainty in Cost-Effectiveness Analysis. *Medical Decision Making* 27:101–111
- 19
- 20
- 21 Lee C, Chertow G, Zenios S (2009) An Empiric Estimate of the Value of Life: Updating the Renal Dialysis Cost-Effectiveness Standard. *Value in Health* 12(1):80–87
- 22
- 23
- 24 Lindhjem H, Navrud S, Braathen N, Biaisque V (2011) Valuing mortality risk reductions from environmental, transport, and health policies: a global meta-analysis of stated preference studies. *Risk Analysis* 31(9):1381–1407
- 25
- 26
- 27 Matell M, Jacoby J (1972) Is There an Optimal Number of Alternatives for Likert-Scale Items? Effects of Testing Time and Scale Properties. *Journal of Applied Psychology* 56(6):506–509
- 28
- 29
- 30 Neumann P, Cohen J, Weinstein M (2014) Updating Cost-Effectiveness—The Curious Resilience of the \$50,000-per-QALY Threshold. *New England Journal of Medicine* 371(9):796–797
- 31
- 32

- 1 Ntzoufras I (2009) Bayesian Modeling Using WinBUGS, 1st edn. John Wiley &  
2 Sons, Inc
- 3 Obenchain R (2008) ICE Preference Maps: Nonlinear Generalizations of Net  
4 Benefit and Acceptability. *Health Services and Outcomes Research Method-*  
5 *ology* 8:31–56
- 6 Orlovsky S (1978) Decision-making with a fuzzy preference relation. *Fuzzy Sets*  
7 *and Systems* 1:155–167
- 8 Ramík J, Římánek J (1985) Inequality relation between fuzzy numbers and its use  
9 in fuzzy optimization. *Fuzzy Sets and Systems* 16:123–138
- 10 Sadatsafavi M, Najafzadeh M, Marra C (2008) Acceptability curves could be  
11 misleading when correlated strategies are compared. *Medical Decision Mak-*  
12 *ing* 28(3):306–307
- 13 Sen A (1970) *Collective Choice and Social Welfare*, 1st edn. Holden Day
- 14 Severens J, Brunenberg D, Fenwick E, O'Brien B, Joore M (2005) Cost-  
15 Effectiveness Acceptability Curves and a Reluctance to Lose. *Pharmacoeco-*  
16 *nomics* 23(12):1207–1214
- 17 Smithson M (1987) *Fuzzy Set Analysis for Behavioral and Social Sciences*, 1st  
18 edn. Springer-Verlag
- 19 Tan-Torres Edejer T, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans D,  
20 Murray C (2003) *Making Choices in Health: WHO Guide to Cost Effectiveness*  
21 *Analysis*, 1st edn. World Health Organization
- 22 Viscusi W, Aldy J (2003) The Value of a Statistical Life: A Critical Review  
23 of Market Estimates Throughout the World. *Journal of Risk and Uncertainty*  
24 27(1):5–76
- 25 Whitehead A (2002) *Meta-Analysis of Controlled Clinical Trials*, 1st edn. John  
26 Wiley & Sons, Ltd
- 27 World Health Organization (2001) *Macroeconomics and Health: Investing in*  
28 *Health for Economic Development. Report of the Commission on Macroecon-*  
29 *omics and Health*, 1st edn. World Health Organization
- 30 Zadeh L (1965) Fuzzy Sets. *Information and Control* 8:338–353
- 31 Zhao J, Kling C (2001) A new explanation for the WTP/WTA disparity. *Eco-*  
32 *nomics Letters* 73:293–300