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

SGH KAE Working Papers Series Number: 2016/006 April 2016

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

Michał Jakubczyk

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

Tippie College of Business, The University of Iowa, USA

Abstract

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

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

JEL: C44; C13; D81; D61

Highlights:

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

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 $\text{WTA}/\text{WTP} \approx 10$ 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, €1≈4.4 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.

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

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

2 Fuzzy willingness-to-pay/accept

2.1 Fuzzy preferences on cost-effectiveness plane

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

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

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

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

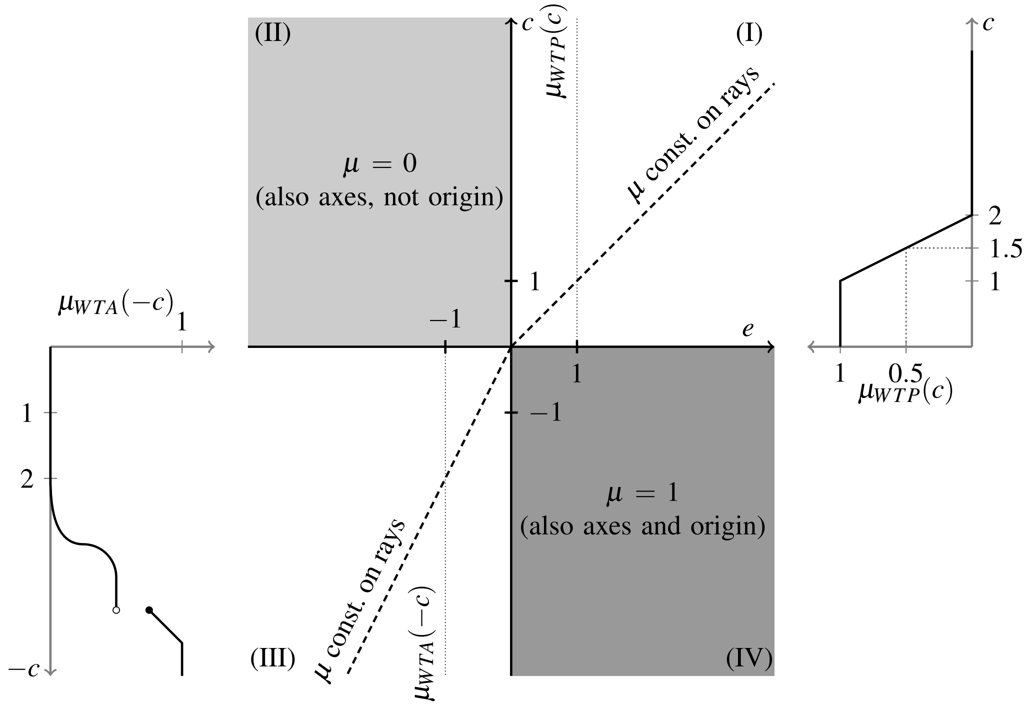

Figure 1: Fuzzy preference relation in cost-effectiveness plane (middle) and its relation to fuzzy willingness-to-pay/accept (right and left, respectively).

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 $\mu_{\text{WTP}}(0) = 1$, $\mu_{\text{WTA}}(0) = 0$, $\mu_{\text{WTP}}(\cdot)$ is non-increasing, $\mu_{\text{WTA}}(\cdot)$ 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.

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

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

We may be tempted to check if WTA>WTP. This requires rephrasing the question in terms of fuzzy approach: we now ask if μ_{WTA} is shifted rightward comparing to μ_{WTP} (apart from a horizontal flip). The WTP-WTA disparity would then mean 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 is strict (resembling the standard fuzzy numbers inequalities, Ramík and Římánek, 1985). It is not obvious how to conduct a statistical comparison (and still, with the survey we obtain Likert answers, not continuous membership). We might compare answers for fWTP vs fWTA (flipped around 3) using Wilcoxon paired test. If fWTP = fWTA, then H_0 is true. Unfortunately, the test rejects H_0 even when fWTA is not shifted, e.g. when fWTA is flatter (options 2–4 used more often). Then, testing individual λs separately would reject H_0 in one direction for small λs, and in the other for large, while the result for the pooled λs depends on the structure of λ in the survey. A different approach is proposed in section 4.

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

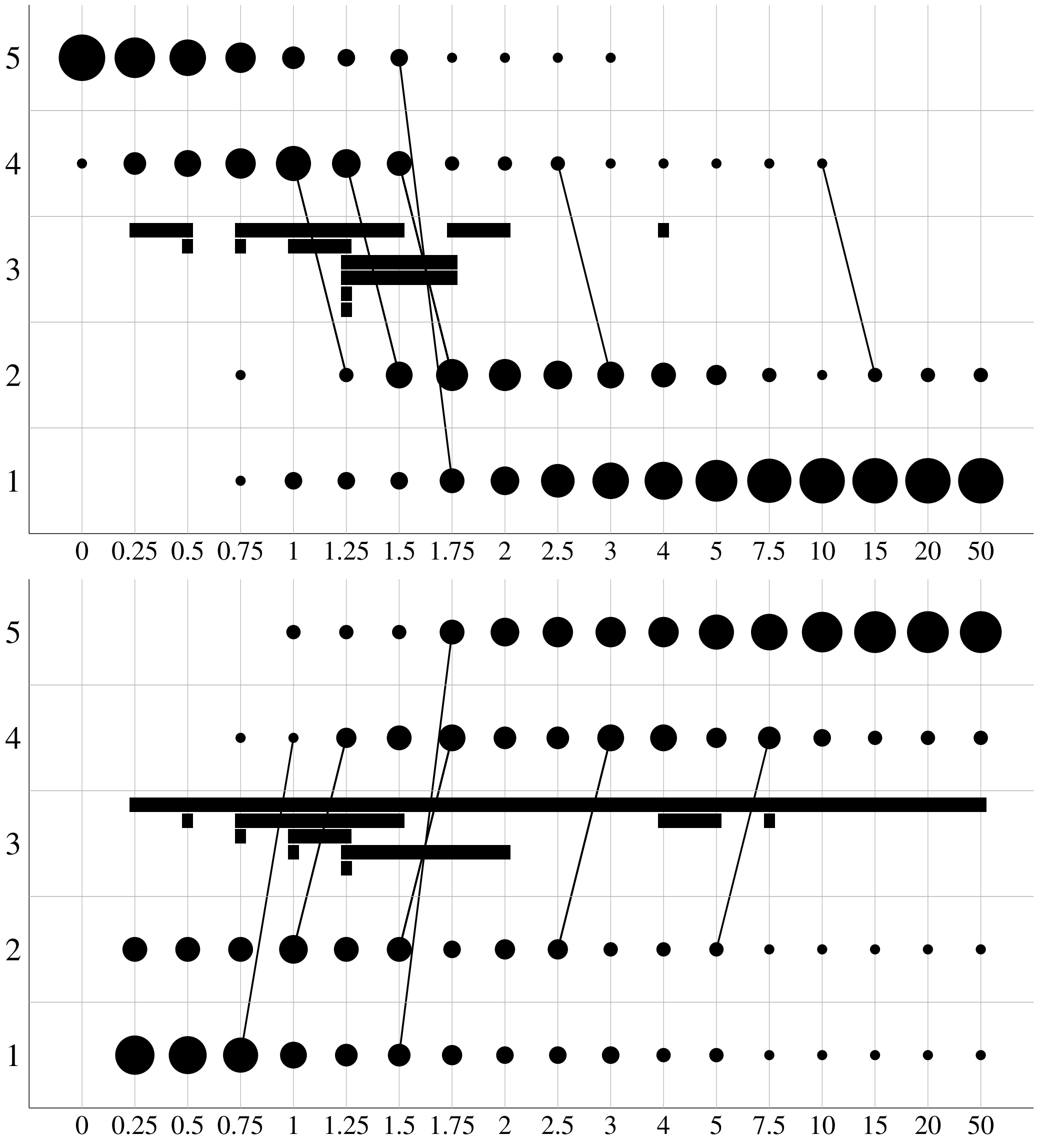

Figure 2: Survey results for WTP/WTA (above/below), values in horizontal axis in hundreds of 000s of PLN/QALY, answers (vertical axis) from a 5-level Likert scale (1—definitely disagree, 5—definitely agree). Horizontal bars represent individuals, circles—the fraction of respondents, lines—jumps across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

and a single value (unfortunately, this wasn't asked for WTA). On average the range amounted to $(88.9;125)$ and the single value to 105 (all results in 000s PLN/QALY). Hence, the freely reported values corresponded to λs towards which the respondents felt quite convicted in Likert-based questions. For each respondent I took the smallest range of λs containing the whole freely reported range, and calculated the average Likert response for these λs (for a respondent reporting 30–90 I consider $\lambda = 25$, 50, 75, and 100). The average (between the respondents) of these means amounted to 3.84, median to 4, 63.6% respondents had a mean ≥ 4, and only 2 respondents (9.1%) <3. The analysis of Likert answers for the single freely reported WTP (if necessary, interpolating for two closest λs) yields, similarly, the average of 3.92 and the median of 4. Thus, using Likert-based questions seems better at assessing the membership function for fWTP/fWTA than relying on directly reported ranges of values.

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

3 Fuzzy choice functions under certainty

3.1 Evaluating decision alternatives with fuzzy net benefit

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

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

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

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

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

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

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

$$A_{\text{F}}(\alpha) = \{x \in \mathbb{R} : \mu_{\text{F}}(x) \geq \alpha\}, \tag{1}$$

for $\alpha \in (0,1]$ and $A_{\text{F}}(0) = \cup_{\alpha \in (0,1]} A_{\text{F}}(\alpha)$. The following useful lemma holds (because we work on *sets* we have to use special addition and product operators).

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

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

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

Two corollaries follow.

Corollary 1. *Assume* (e_2, c_2) *is Pareto-dominated by* (e_1, c_1)*, i.e.* $e_2 \leq e_1 \wedge c_2 \geq c_1$ *(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, if* $c_2 > c_1$ *or (*$e_2 < e_1$ *and* $\mu_{fWTP}(x) > 0$*,* $\mu_{fWTA}(x) > 0$ *for some* $x > 0$*), then* $\exists_{\alpha \in (0,1]} : A_{fNB(e_2,c_2)} \neq A_{fNB(e_1,c_1)}$*.*

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

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) *is Pareto-dominated by some* $\gamma(e_1, c_1) + (1 - \gamma)(e_2, c_2)$ *with* $\gamma \in [0, 1]$*, and either* $e_1, e_2, e_3 \geq 0$ *or* ≤ 0*. Then* $\forall_{\alpha \in (0,1]} : A_{fNB(e_3,c_3)} \subset \left(A_{fNB(e_1,c_1)} \cup A_{fNB(e_2,c_2)}\right)$*.*

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

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

3.2 Making a choice

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

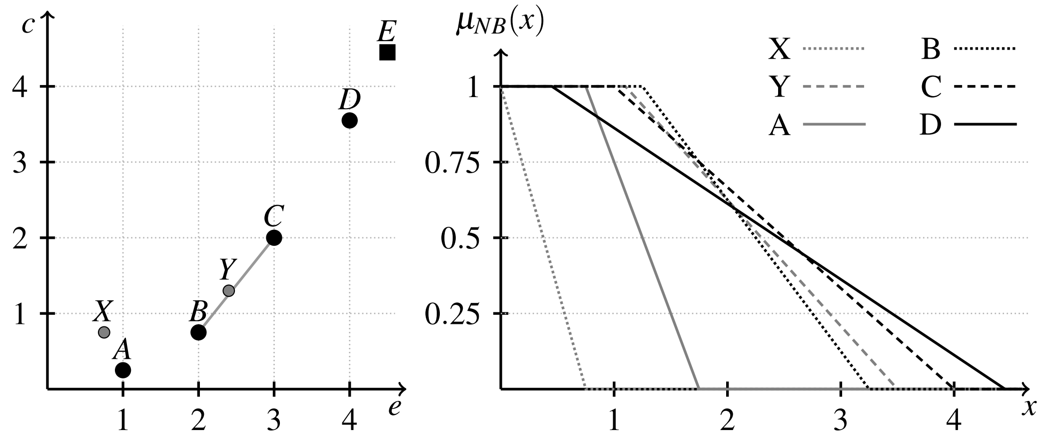

Figure 3: Exemplary fNB membership functions, when μ_{fWTP} decreases linearly in $[1;2]$ (as in Fig. 1). $X = (0.75, 0.75)$, $Y = (2.4, 1.3)$, $A = (1, 0.25)$, $B = (2, 0.75)$, $C = (3, 2)$, $D = (4, 3.55)$, and $E = (4.5, 4.45)$, fNB for E not drawn for clarity.

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 $\mathrm{fNB}(D_2 - D_1) \geq 0$, hence, $\mu_{\mathrm{fNB}(D_2-D_1)}(0)$. 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 $\mathrm{fNB}(D_2 - D_1) \geq 0$ and not $\mathrm{fNB}(D_3 - D_2) > 0$, which now requires selecting the fuzzy logical operators (AND, NOT). I take NOT $\mathrm{fNB}(D_3 - D_2) > 0 = 1 - \mu_{\mathrm{fNB}(D_3-D_2)}(0)$ (typical approach), and the bounded sum AND (see, e.g. Smithson, 1987) to get

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

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

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

where $[P] = 1$ if P is true, and 0 otherwise. Several options may have $\beta > 0$, as the decision maker is not fully confident of her WTP/WTA. Using this approach to make a final decision it would be natural to select $\arg\max_{i\in I}\beta_i$. In the example in 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 E for now). This approach is the most fuzzy one: it postpones the crispification until the last possible moment, just when the crisp choice is being made.

More technically, βs can be calculated (i.e. integrals in equation 2 are well defined): Lemma 1 and monotonicity of μ_{fNB} (from the monotonicity of μ_{fWTP}) imply that the integrand will be equal to 1 over a single segment of αs. The situation gets complicated when considering the whole CE-plane. No so intuitive derivation can be presented (still, the method and its graphical interpretation appeals to intuition). With no additional stipulations regarding μ_{fWTP} and μ_{fWTA} the resulting μ_{fNB} for various alternatives can intersect many times (countably many at maximum, though) for various αs, and so we would have to add up the lengths of several segments in equation 2. This is unlikely to cause any problems in real applications (μ_{fWTA} and μ_{fWTA} would be approximated by regular functions) and is not pursued here.

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

3.2.2 Average fuzzy net benefit

The remaining two choice functions are based on crispifying fNB, and comparing these crisp representations. First, I continue with measuring the vertical changes in μ_{fNB}, but I interpret them as probabilities (conveniently summing to 1; the membership function is treated as a, perhaps flipped horizontally, cumulative distribution function). Then for each D_i, take

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

and calculate an average fNB (ANB): $\text{ANB}_i := \int_0^1 \tau_i(\alpha)\,\mathrm{d}\alpha$, i.e. average out the bounds of α-cuts. Then the choice is simply $\arg\max_{i\in I}\text{AfNB}_i$. Technically, the integral exists, as τ is non-decreasing and bounded (for a given (e,c)). Considering the complete CE-plane does not change the intuition behind the derivation nor the feasibility to use. The method preserves the Chernoff property: the evaluation of each alternative is independent of other options. The obvious disadvantage is, again, the necessity to interpret the membership function as an interval scale.

3.2.3 Median fuzzy net benefit

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

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

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

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

Proposition 1. *Take n decision alternatives, D_i. If D_{i^*} maximizes MNB, then fNB_{i^*} weakly includes fNB_i for any i not maximizing MNB, i.e.*[1]

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

and fNB_i weakly includes fNB_{i^*} at maximum to the same degree:

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

Moreover, two implications hold:

- *if μ_{fWTP} and μ_{fWTA} are strictly decreasing (for values within $(0,1)$ interval), then the above inequalities are strict;*

[1] With the following intuition. Consider crisp sets, A, B, subsets of some universe Ω. 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).

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

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

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

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

4 Calculating the 0.5-cut for fuzzy WTP & WTA

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

4.1 Hypothesis testing

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

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

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

4.2 Data transformation

Above I analysed each λ separately, but looked at the respondents' jointly. In two remaining approaches I proceed conversely: I consider each respondent individually, looking at all the λs for which the middle Likert option was chosen (interpreted as $\mu(\lambda) = 0.5$) simultaneously. I call this range of λs an *indecisiveness range* (IR), and will use IR to estimate a single IP value.

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

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

The assumptions suffice to calculate IRs for WTP. In case of WTA, however, two respondents used only options 1 & 2, and one respondent only option 3, for all the λs, thwarting the calculation of IR. I removed all three from the sample, based on two reasons. 1) These respondents do not conform to the *criteria tradeability* axiom of J&K: they seem to, in principle, disagree that the decision maker should sometimes sacrifice effectiveness to make savings. The decision support methods developed in the present paper accept such trade-offs (and aim to express them quantitatively), and should not be based on the opinions in such a fundamental disagreement with the foundations. 2) We aim here mostly to illustrate the estimation methods, and not to come up with ultimate, ready-to-use estimates. The latter would demand a larger sample and probably using more λs in the questionnaire (or a possibility to freely report large values if the scale is insufficient). Still, a further research is needed to consider how this, effectively infinite, WTA should be accounted for quantitatively (how to combine finite and infinite WTA in a mathematical framework).

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

4.3 Hierarchical Bayesian modelling

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

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

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

I used non-informative priors (normal for μ, gamma for σ^2 and λ) and estimated the model with MCMC in JAGS/R (10,000 burn-in iterations, 50,000 of actual iterations, thinning 5). The mean of the posterior was taken as the estimate, and percentiles 2.5% and 97.5% as boundaries of the 95% credible interval (CI).

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

4.4 A meta-analytic approach with bootstrap

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

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

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

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

4.5 Comparison of approaches

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

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

The last two methods require the middle option, so as to account for possibly wide IR (otherwise we would underestimate the intra-respondent uncertainty regarding the location of IP). The middle option cannot, however, be too inclusive (e.g. *neither entirely agree, nor entirely disagree* in a 3-level Likert) for the last two methods, as that would change the estimand. The hypothesis testing approach can be used irrespectively whether or not the middle level is used, how it is worded (e.g. *neither/nor* or *I don't know*, as long as it is symmetric, i.e. not leaning towards agreeing or disagreeing), and whether it is framed to be more attractive for respondents. Still, making the middle answer too attractive reduces the power of the test, as fewer observations constitute the actual sample for the Dykstra et al (1995) test. Matell and Jacoby (1972) showed that using more (odd number of) levels decreases the frequency of selecting the middle option—hence. All the methods might profit from using a greater than five, odd number of levels. Even though the levels might then lose natural interpretation (section 2.2), if respondents symmetrically behave on two sides of the middle option that would increase the precision of IP estimation.

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

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

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

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

Finally, we might try to explicitly allow for $\mu_{\text{fWTP}}(x) = 0.5$ ($\mu_{\text{fWTA}}(x) = 0.5$) for a range of x. Looking at the definition of MfNB we now need to find the largest such x for fWTP and the smallest for fWTA. We don't need to change anything in the hypothesis testing approach, and then, conveniently, we infer that this value amounts to 150 (000s PLN/QALY) for both WTP and WTA. For the remaining two approaches we would be interested in the respective ends of the IR, u_i or l_i. It is difficult to naturally quantify the individual-level error of u_i (l_i), and so the most natural approach is to average the observed values, and account for the sampling error with, e.g. bootstrap over the respondents. We then obtain (in 000s PLN/QALY) 165.96 with 95%CI $= (131.21; 219.49)$ for WTP, and 140.53 with 95%CI $= (103.84; 197.09)$ for WTA. However numerically greater WTP is WTA, the difference is not statistically significant (even a 90%CI for difference contains 0). Hence, whatever approach we take, there seems to be no rationale to systematically differentiate WTP from WTA.

5 Uncertainty & sensitivity analysis

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

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

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

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

In HTA there is stochastic uncertainty, and SA is used to illustrate its impact. The main source is that (e_i, c_i) are estimated based on clinical trials, meta-analyses, or pharmacoeconomic models (cf. Briggs et al, 2012). The current framework, adds two elements. Firstly, the IPs are also estimated, and so using MNB removes fuzziness in the end, introducing more stochastic uncertainty. Secondly, the selection of baseline technology (with respect to which (e_i, c_i) are calculated) impacts the results by moving the points between halves of the CE-plane, important when fWTP and fWTA differ. Often a mix of technologies will be used. Then we can calculate (e_i, c_i) with different baseline technologies in this mix, with proper market shares taken as weights/probabilities. This uncertainty can be simply joint with (e, c)-estimation uncertainty (but will be left out, for clarity, in the examples that follow).

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

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

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

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

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

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

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

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

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

It is most informative to calculate some low percentiles of MNB distributions, to see how risky the available alternatives are. For example, the 5% percentile is equal to -0.47, -0.23, and -0.35 (for D_1–D_3), and so the risk averse decision maker should favour the null option. Analysing the percentiles of MNB has a nice property that moving away from the origin in CE-plane (to be more precise: increasing the absolute value of *e*) increases the risk. That seems intuitive, as implies that larger trade-offs are being made, which should feel risky for a decision maker uncertain of own WTP/WTA. That is a new property, absent in standard CEAC analysis (and criticized, e.g. by Koerkamp et al, 2007).

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

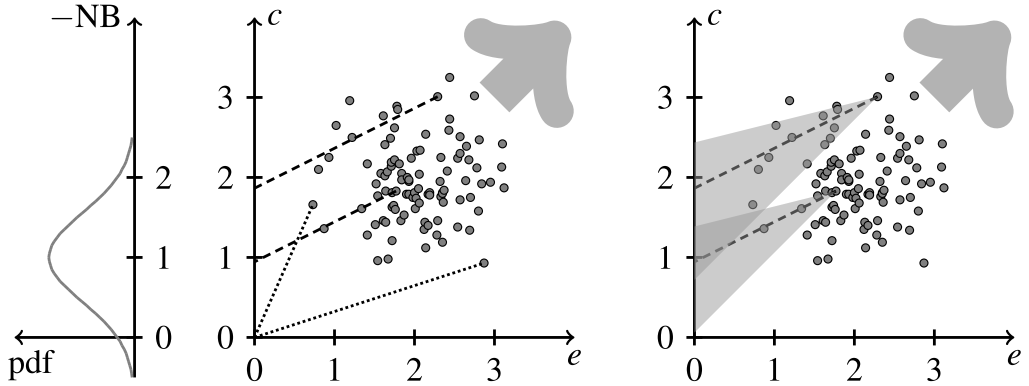

Figure 4: Increasing the absolute value of effect does not change the total uncertainty of NB when WTP is given (in central panel, left panel represents the density function of NB) and increases it when WTP is estimated with uncertainty (right panel).

As mentioned above, the probability of maximizing MNB could be (almost) read off the standard CEAC, by averaging over the range of WTPs. That is not the case for expected value of perfect information (EVPI). An example below illustrates that the uncertainty in IPs for WTP and WTA results in qualitatively new phenomena, that cannot be seen in standard graphs used to illustrate EVPI. Of course the very amount of uncertainty is larger, and so EVPI increases, but the difference is also a qualitative one. Consider comparing $(e,c) = (k,k)$ with SD equal to 1 (for e and c) for $k = 1,2,5,10,20$ versus $(0,0)$, i.e. consider a set of five decision problems. Larger k represents larger shift outwards. In Figure 5 the EVPI are presented for various WTP. As k increases (represented by a darker shade), EVPI seems to decrease for all but one value of WTP, giving the impression that overall there is less uncertainty in the problem. When we assume WTP is only given as a distribution (here, a uniform $[1/2, 3/2]$), then the resulting EVPI (now, a single number) increases with k: 0.59, 0.61, 0.8, 1.36, and 2.55.

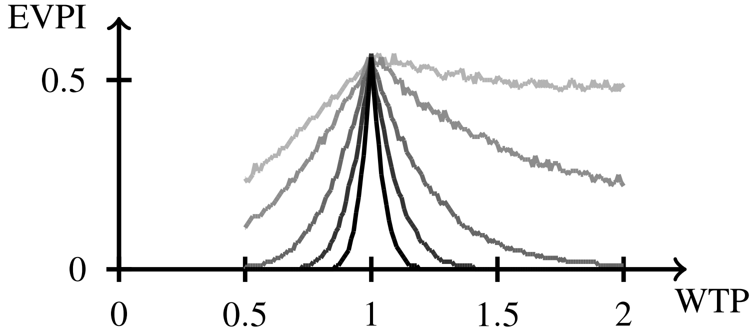

Figure 5: EVPI for a comparison of (e, c) *distributed around* (k, k) *vs* $(0, 0)$ *when* $k = 1, 2, 5, 10, 20$ *(darker shade, larger k).*

6 Concluding remarks

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

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

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

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

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

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

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

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

Acknowledgements

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

Proofs

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

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

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

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

$\square$

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 $0 \in A_{\text{fNB}(\Delta_e, \Delta_c)}(\alpha)$ for any $\alpha \in (0, 1]$; then use $\oplus$. Strict version follows trivially. If (e_1, c_1), (e_2, c_2) are separated by y-axis, then use $(0, {}^{(c_1+c_2)}/_2)$ as an intermediary. $\square$

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

Proposition 1. Proving the first part. Take any $x \in (\text{MNB}_i, \text{MNB}_{i^*})$ (the interval is non-empty), $\mu_{\text{fNB}(i)}(x) \leq {}^1/_2 \leq \mu_{\text{fNB}(i^*)}(x)$; using the monotonicity of μ_{fNB} (for i^*, i) yields the result. Proving the first bullet implication: take any $x \in (\text{MNB}_i, \text{MNB}_{i^*})$ (again, exists), $\mu_{\text{fNB}(i)}(x) < {}^1/_2 < \mu_{\text{fNB}(i^*)}(x)$, and use monotonicity again. Proving the last bullet. First consider $e_{i^*} \neq 0 \neq e_{i^{**}}$, and so μ_{fNB} are continuous for i^*, i^{**}. Then $\mu_{\text{fNB}(i^*)}(x) = \mu_{\text{fNB}(i^{**})}(x) = {}^1/_2$, and the rest follows from monotonicity. Now consider $e_{i^*} = 0 = e_{i^{**}}$, then fNBs are equal, crisp numbers (with upper semi-continuous, step membership functions, jumping from 1 to 0), and so weakly include each other to the degree 1. Finally consider $e_{i^*} \neq 0 = e_{i^{**}}$. $\mu_{\text{fNB}(i^*)}$ is continuous and monotonic, and $\text{fNB}_{i^{**}}$ is a crisp number. It easily follows (considering $x = \text{MNB}_{i^*}$) that fNB_{i^*} weakly includes $\text{fNB}_{i^{**}}$ to the degree ${}^1/_2$. Approaching this x from right yields the weak inclusion to the same degree. $\square$

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

Proposition 3. Looking at the definitions of fNB (Def. 1), α-cuts (eq. 1), and τ (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 the continuous mapping theorem (cf. Billingsley, 1999) we get the result. $\square$

References

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

Araki D, Kamae I (2015) The Augmented Representation of the Cost-effectiveness Acceptability Curve for Economic Evaluation of Health Technology. Kobe Journal of Medical Sciences 61(1):E9–E18

Bellavance F, Dionne G, Lebeau M (2009) The value of a statistical life: a meta-analysis with a mixed effects regression model. Journal of Health Economics 28(2):444–464

Billingsley P (1999) Convergence of Probability Measures, 2nd edn. Wiley

Briggs A, Weinstein M, Fenwick E, Karnon J, Sculpher M, Paltiel A, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force (2012) Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value in Health 15:835–842

Doucouliagos C, TD S, Giles M (2012) Are estimates of the value of a statistical life exaggerated? Journal of Health Economics 31(1):197–206

Dubois D, Prade H (1980) Fuzzy Sets and Systems: Theory and Applications. New York: Academic Press

Dubourg W, Jones-Lee M, Loomes G (1994) Imprecise Preferences and the WTP-WTA Disparity. Journal of Risk and Uncertainty 9:115–133

Dykstra R, Kochar S, Robertson T (1995) Likelihood ratio tests for symmetry against one-sided alternatives. Annals of the Institute of Statistical Mathematics 47(4):719–730

Efron B, Tibshirani R (1993) An Introduction to the Bootstrap, 1st edn. Chapman and Hall/CRC

Fenwick E, Claxton K, Sculpher M (2001) Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 10(8):779–787

Fenwick E, O'Brien B, Briggs A (2004) Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. Health Economics 13:405–415

Garber A (2000) Advances in Cost-Effectiveness Analysis of Health Interventions. In: AJ Culyer JN (ed) Handbook of Health Economics. Volume 1A, North-Holland, pp 181–221

Grosse S (2008) Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Review of Pharmacoeconomics & Outcomes Research 8(2):165–178

Horowitz J, McConell K (2002) A Review of WTP/WTA Studies. Journal of Environmental Economics and Management 44(3):426–447

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

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

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

Jakubczyk M (2016) Using a fuzzy approach in multi-criteria decision making with multiple alternatives in health care. Multiple Criteria Decision Making forthcoming:

Jakubczyk M, Kamiński B (2010) Cost-Effectiveness Acceptability Curves—Caveats Quantified. Health Economics 19:955–963

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

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

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

Lindhjem H, Navrud S, Braathen N, Biausque 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

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

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

Ntzoufras I (2009) Bayesian Modeling Using WinBUGS, 1st edn. John Wiley & Sons, Inc

Obenchain R (2008) ICE Preference Maps: Nonlinear Generalizations of Net Benefit and Acceptability. Health Services and Outcomes Research Methodology 8:31–56

Orlovsky S (1978) Decision-making with a fuzzy preference relation. Fuzzy Sets and Systems 1:155–167

Ramík J, Římánek J (1985) Inequality relation between fuzzy numbers and its use in fuzzy optimization. Fuzzy Sets and Systems 16:123–138

Sadatsafavi M, Najafzadeh M, Marra C (2008) Acceptability curves could be misleading when correlated strategies are compared. Medical Decision Making 28(3):306–307

Sen A (1970) Collective Choice and Social Welfare, 1st edn. Holden Day

Severens J, Brunenberg D, Fenwick E, O'Brien B, Joore M (2005) Cost-Effectiveness Acceptability Curves and a Reluctance to Lose. Pharmacoeconomics 23(12):1207–1214

Smithson M (1987) Fuzzy Set Analysis for Behavioral and Social Sciences, 1st edn. Springer-Verlag

Tan-Torres Edejer T, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans D, Murray C (2003) Making Choices in Health: WHO Guide to Cost Effectiveness Analysis, 1st edn. World Health Organization

Viscusi W, Aldy J (2003) The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World. Journal of Risk and Uncertainty 27(1):5–76

Whitehead A (2002) Meta-Analysis of Controlled Clinical Trials, 1st edn. John Wiley & Sons, Ltd

World Health Organization (2001) Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics and Health, 1st edn. World Health Organization

Zadeh L (1965) Fuzzy Sets. Information and Control 8:338–353

Zhao J, Kling C (2001) A new explanation for the WTP/WTA disparity. Economics Letters 73:293–300