M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G

Vol. 12 2017

Bogumił Kamiński[*]
Michał Jakubczyk[**]

COMPARING THE CRISP AND FUZZY APPROACHES TO MODELLING PREFERENCES TOWARDS HEALTH STATES

DOI: 10.22367/mcdm.2017.12.06

Abstract

Understanding societal preferences towards health is vital in public decisions on financing health technologies. Thought experiments in which respondents choose between health states are used to understand the importance of individual criteria. Competing models of preference structure can be compared by their ability to explain empirical observations. One of the key challenges when constructing such models is that they have to aggregate preferences defined in multiple-criteria space. In the present paper, we test whether treating the impact of health worsening (defined using EQ-5D-5L descriptive system, i.e. decomposing health status in five criteria) as a fuzzy concept can improve the model fit. To test if fuzzy approach to multiple-criteria preferences aggregation is valid, we compare a standard, crisp model (SM) with two models using fuzzy sets (JKL, previously proposed in the literature; and FMN introduced here). We find FMN better than SM, and SM better than JKL. Anxiety/depression and pain/discomfort seem to weigh most in preferences. According to FMN, self-care and usual activities are associated with largest imprecision in preferences. The respondents are susceptible to framing effects when time unit is changed: e.g. measuring the duration in days results in short intervals mattering more than when expressed as weeks. We conclude that the fuzzy-based framework is promising, but requires careful work on the exact specification.

[*] SGH Warsaw School of Economics, Decision Analysis and Support Unit, Warsaw, Poland, e-mail:bogumil.kaminski@sgh.waw.pl.

[**] SGH Warsaw School of Economics, Decision Analysis and Support Unit, Warsaw, Poland, e-mail:michal.jakubczyk@sgh.waw.pl.

Keywords: fuzzy modelling, discrete choice experiment, health-related quality of life, preference elicitation.

1 Introduction

Understanding people's preferences towards health is important; it can serve in health technology assessment (HTA) to evaluate benefits of increasing life expectancy or improving health-related quality of life (HRQoL). Hence, the elicitation of preferences serves a prescriptive purpose: to suggest a course of action to be taken, typically, by the public regulator on behalf of the society.

There are two steps in the preference-elicitation process. First, the mathematical representation of the preferences is constructed; then, its parameters are estimated based on empirical data. Regarding the former, typically a quality-adjusted life years (QALY) model is used in HTA (Weinstein et al., 2009): a health state Q is assigned a number, $u(Q)$, with the interpretation that $T \times u(Q)$ denotes the von Neumann-Morgenstern utility of spending T years in Q, where $u(\text{dead}) = 0$ and $u(\text{full health}) = 1$ (see Bleichrodt et al., 1997; Miyamoto et al., 1998). The key challenge is that Q is usually evaluated using multiple criteria, and u is a function aggregating them into a single instrument that can be used operationally. In this text, we use the EQ-5L-5D system to define HRQoL, which implies a five-attribute description of Q.

Regarding the parameter estimation, usually either a time trade-off (TTO) or a discrete choice experiment (DCE) is used to collect data on preferences. In TTO, we attempt to determine the time T, such that T years in full health is equivalent to 10 years in Q; in DCE, the respondent faces a series of pairwise comparisons between two states, Q_1 and Q_2, lasting for T_1 and T_2 (in *DCE with duration*), respectively (immediate death may also be used). The results of TTO or DCE, combined with the QALY model assumptions (and the resulting econometric models), can serve to assign utilities to health states.

However popular in applied HTA the QALY model is, its founding assumptions are often criticized (e.g. Attema et al., 2010; Pettitt et al., 2016; Beresniak et al., 2015). Among various lines of critique, Jakubczyk and Kamiński (2017) and Jakubczyk (2015) suggested that fuzzy sets (a concept introduced by Zadeh, 1965) can be used to define preferences towards health states (in the context of health vs money trade-offs); the approach is motivated by the observation that a lack of market experience can lead to an inherent imprecision in preferences. In the present paper, we aim to compare the standard (crisp) approach with a fuzzy-based one. Even though there is no descriptive motive in the health preference research – we do not strive to predict somebody's choices (as people rarely actually choose between health states) – the model fit seems a natural way to evaluate the credibility of the elicited values (Jakubczyk et al., 2017).

We compare three approaches: 1) a standard QALY-model-based, crisp model, as a benchmark; 2) a fuzzy-based approach proposed by Jakubczyk et al. (2017) (JKL, henceforth); 3) an alternative fuzzy-based specification, developed in the present paper (FMN, henceforth). We use the model fit as a basic measure of model quality (minus log likelihood), but also discuss the face validity of estimated parameters.

In the next section, we first present more details on how health states are defined and the specifications of all three approaches. In section 3, we introduce the dataset and the numerical approach used to estimate the parameters of the models. In section 4, we present the results, compare the models with respect to the insight on the impact of health on utility, and comment on the predictive validity of the approaches. Finally, we discuss our findings.

2 Fuzzy modelling of preferences towards health states

2.1 Benchmark, crisp model

Health states are often described with the EQ-5D-3L (or 5L) descriptive system (Brooks et al., 1996; Herdman et al., 2011), i.e. using five dimensions (or *criteria* in decision modelling parlance): mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). In each dimension, health can be at one of three (in 3L) or five (5L) levels, denoting no problems (level 1) or more and more severe problems (consecutive levels). In such a descriptive system, a health state is denoted by five consecutive digits; in particular, 11111 denotes full health (FH), and 55555 (we focus on 5L case henceforth) denotes the worst (in the descriptive system considered) possible health state[1].

There are 3125 health states in the EQ-5D-5L descriptive system, making it virtually impossible to elicit the utility for all of them. For this reason, a model is fitted to the data collected for a subset of states, and then the utilities of all the states can be approximated via extrapolation. Typically, the utility of a health state Q is calculated relative to the utility $u(\mathrm{FH}) = 1$, in the form:

$$u(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} \alpha_{i,j} d_{i,j}(Q), \tag{1}$$

where

[1] This notation is in standard use in the literature (and, e.g., not a vector-like $(5, 5, 5, 5, 5)$); hence, we use it here.

- i indexes the dimensions,
- j indexes the levels (no disutility for level 1; hence, omitted in the formula above),
- $d_{i,j}$ is a dummy denoting whether dimension i is at level j,
- parameters $\alpha_{i,j}$ represent the preference structure (again, no disutility at level 1).

We expect that $\alpha_{i,j}$ is positive and increasing with j. Often a constant term, α_0 is added (if Q differs from FH, i.e. if at least one $d_{i,j} = 1$). Because JKL did not use it in their specification and because α_0 is difficult to interpret, we omit it in the basic benchmark specification here[2]. Nonetheless, we also present the version with the constant term, differing little in terms of the model fit.

When the above model is estimated based on TTO data, an error term, ε, is added to eq. (1) (otherwise, no set of parameters α could fit the observed data). In DCE, when two health states, Q_A and Q_B, considered for T_A and T_B years, respectively, are compared, it is often assumed that the probability of Q_A being selected is given by an exponential version of the Bradley-Terry approach (e.g. Bansback et al., 2012):

$$P(Q_A, T_A, Q_B, T_B) = \frac{\exp\left(u(Q_A) \times T_A\right)}{\exp\left(u(Q_A) \times T_A\right) + \exp\left(u(Q_B) \times T_B\right)}. \quad (2)$$

In the benchmark model, we use eq. (2) to estimate the parameters of eq. (1).

Notice that equation (2) allows the utility to be negative (which is not a problem thanks to the exponential function), and indeed some health states are perceived as worse than dead. Immediate death is equivalent to 0 utility (i.e. is equivalent to any state with duration zero). Jakubczyk et al. (2017) point out that in the above formula there is always a positive probability of any state being selected, however worse it is (even dominated) than the other one.

In equation (2), the rescaling of T_A and T_B (i.e. multiplying both by the same positive constant) changes the probability. Therefore, whether the time is expressed as years (usually the case) or months, weeks, days, should also impact the result. At the same time, presenting the choice in the specific time unit might frame the problem differently (e.g. the subjective perception of *six months* might differ from *half a year*). In our dataset, we use various time units (days, weeks, months, and years). Hence, we introduce three parameters,

[2] The constant term was interpreted in terms of dimensions complementarity by Jakubczyk (2009); it could be also interpreted to reflect the fact that the state 11111 might also include some minor health problems.

τ_1, τ_2, and τ_3, scaling the duration when days, weeks, or months are used, respectively. For example, if time is expressed in days in the task, then the formula for probability becomes:

$$P(Q_A, T_A, Q_B, T_B) = \frac{\exp\left(u(Q_A) \times T_A \times \tau_1\right)}{\exp\left(u(Q_A) \times T_A \times \tau_1\right) + \exp\left(u(Q_B) \times T_B \times \tau_1\right)}. \quad (3)$$

The parameters to be estimated are α (20) and τ (3). We expect τ_i to be < 1 and increasing with i. If there is no framing effect of a time unit, we should get $\tau_1 = 1/365$, $\tau_2 = 1/52$, and $\tau_3 = 1/12$.

2.2 JKL fuzzy model

We only briefly reintroduce the JKL model (Jakubczyk et al., 2017), and the reader is encouraged to see the original publication for details. JKL suggested to treat the utility of being in state Q for T years, $u(Q, T)$, as a fuzzy set. For simplicity, they used a piecewise linear membership function, $\mu_{u(Q,T)}(x)$; $\mu_{u(Q,T)}(x) = 1$ for low values ($x \leq L(Q) \times T$), $\mu_{u(Q,T)}(x) = 0$ for high values ($x \geq H(Q) \times T$), where $L(Q)$ and $H(Q)$ are parameters characterizing health state Q. Then, $\mu_{u(Q,T)}(x)$ is linearly decreasing between $L(Q) \times T$ and $H(Q) \times$ $\times$ T (or jumping discontinuously, if $L(Q) \times T = H(Q) \times T$). JKL interpret $\mu_{u(Q,T)}(x)$ as the conviction that being in Q for T years gives the utility of at least x.

Analogously to eq. (1), $L(Q)$ and $H(Q)$ are given as linear combinations of dummies for dimensions and levels defining Q:

$$L(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} h_{i,j} d_{i,j}(Q), \quad (4)$$

$$H(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} l_{i,j} d_{i,j}(Q). \quad (5)$$

Parameters l and h define the range of disutility for a given dimension/level. The larger they are, the bigger the impact of a given health worsening on utility. The more they differ, the larger the imprecision in the perception of disutility. Because of the subtraction, parameters h are used to define $L(Q)$, and parameters l are used to define $H(Q)$.

In the JKL model, two health states are compared in the following way (e.g. in a DCE experiment). The advantage of (Q_1, T_1) over (Q_2, T_2), namely $\delta_{(Q_1,T_1),(Q_2,T_2)}$, is given as:

$$\sup_{x \in \mathbb{R}} \left(\mu_{u(Q_1,T_1)}(x) - \mu_{u(Q_2,T_2)}(x)\right),$$

and the advantage the other way round is defined analogously. The parameters δ must be in the $[0, 1]$ range. It can happen that both δs are positive. Then, the net advantage of advantage of (Q_1, T_1) over (Q_2, T_2), $\Delta_{(Q_1,T_1),(Q_2,T_2)}$, is given as $\delta_{(Q_1,T_1),(Q_2,T_2)} - \delta_{(Q_2,T_2),(Q_1,T_1)}$, and the resulting $\Delta \in [-1, 1]$.

The probability of (Q_1, T_1) being chosen instead of (Q_2, T_2) is given as a function of the net advantage: $P(\Delta) = \frac{(\Delta+1)^\rho}{2}$, for $\Delta \leq 0$, with a non-negative parameter ρ to be estimated (and for $\Delta > 0$, the probability is calculated using the assumption that $P(\Delta) + P(-\Delta) = 1$). The value $\rho = 1$ leads to Δ being transformed linearly into probability, $\rho < 1$ results in probability remaining at around 50% for many values of Δ, and $\rho > 1$ results in a probability being sensitive to Δ values differing even slightly from zero.

Let us notice the following features of JKL's model. First, multiplying T_1 and T_2 by the same strictly positive number does not change the preferences (parameters δ, Δ, and $P(\cdot)$). Hence, no counterparts of τ are needed (as long as the same time unit is used in both states compared). Second, for large enough differences between (Q_1, T_1) and (Q_2, T_2), the probability of one being chosen is equal to 1 (not only approaches 1). Third, this model compares the two health profiles in the conviction space (i.e. the values of membership functions are compared), rather than in the utility space. This last property motivates trying another fuzzy-based approach, presented in the next subsection (in which the first two properties do not hold; hence, the model is more flexible).

2.3 Fuzzy model – a new specification (FMN)

Again, the utility of living in Q for T years, $u(Q, T)$, is defined by two numbers, $L(Q) \times T$ and $H(Q) \times T$. The values $L(Q)$ and $H(Q)$ are defined as in eq. (4). In the present specification, though, the membership function, $\mu_{u(Q,T)}(x)$ is equal to 1 for $L(Q) \times T \leq x \leq H(Q) \times T$, and 0 otherwise. The interpretation is that the decision maker agrees fully that T years in state Q may correspond to the utility of x, or – putting it differently – cannot rule out x as the utility of (Q, T) and totally rules out any value below $L(Q) \times T$ or above $H(Q) \times T$.

We then define the advantage of one profile, (Q_1, T_1), over another, (Q_2, T_2), by comparing the middles of 1-cuts of $u(Q, T)$:

$$\Delta_{(Q_1,T_1),(Q_2,T_2)} = \left(T_2 \frac{H(Q_2) + L(Q_2)}{2} - T_1 \frac{H(Q_1) + L(Q_1)}{2} \right) \times \tau_i. \qquad (6)$$

The parameter τ_i is the scaling factor, and i changes with the time unit used to measure T_1 and T_2 (the same time unit assumed): $i = 1$ for days, $i = 2$ for weeks, $i = 3$ for months, and $i = 4$ for years. We normalize $\tau_4 = 1$, and estimate τ_1, τ_2, and τ_3.

We define the ancillary score, π, (subscripts $(Q_1, T_1), (Q_2, T_2)$ suppressed; cf. eq. (2)),

$$\pi = \frac{1}{1 + \exp(-\Delta)}. \tag{7}$$

In this way, we transform Δ in the $\pi \in [0, 1]$ interval, to facilitate interpreting the advantage in terms of probabilities.

In this model, we try to account for the fact that a larger difference between $L(Q)$ and $H(Q)$ denotes a larger imprecision in how the utility of Q is perceived. We want to include the possibility that larger imprecision may dilute the preferences, i.e. shift the probability of one alternative being chosen towards 50%. Specifically, we take:

$$\Theta_{(Q_1,T_1),(Q_2,T_2)} = \left(T_2 \frac{H(Q_2) - L(Q_2)}{2} + T_1 \frac{H(Q_1) - L(Q_1)}{2} \right) \times \tau_i. \tag{8}$$

In the present specification, we use the same vector $[\tau_1, \tau_2, \tau_3]$ when calculating Δ and θ, assuming the time unit is perceived identically in both aspects.

Finally, we define the resulting probability of (Q_1, T_1) being chosen as:

$$P = \frac{\pi - 0.5}{1 + \omega\Theta} + 0.5, \tag{9}$$

where ω is a parameter to be estimated. For $\omega = 0$ there is no impact of imprecision on preferences (and the model can be reduced to a crisp version with $L(Q) = H(Q)$ for all the states).

Summing up, in the FMN specification, we have to estimate parameters h and l (40), ω (one parameter), and τ (3).

3 Methods

3.1 Dataset

We used the data from the DCE predictive competition organised by the International Academy for Health Preference Research (IAHPR), sponsored by The EuroQol Group, available for general public, and described on the IAHPR website (`http://iahpr.org/eq-dce-competition/`, as of 16 Nov, 2016). More details can be found in Jakubczyk et al. (2017).

There were responses from 4074 US respondents, each choosing between two health states in 20 pairwise comparisons. Each health state was described with the EQ-5D-5L descriptive system, and the duration was given (four time units were used: days, weeks, months, and years; the same unit for both states in each pair).

In the modelling, only the aggregated proportion of a given answer is used, not the entire individual answer.

3.2 The approach to estimation

In the data set, there were 1560 different combinations of compared states and times. For each combination c_i, the number of observations, n_i, and the number of choices of option 1, k_i, were recorded.

To estimate the parameters of the model (in all three specifications), we employed the maximum likelihood estimation:

$$\max \quad \mathcal{L} = \sum_{i=1}^{1560} k_i \ln(\Pr(c_i)) + (n_i - k_i) \ln(1 - \Pr(c_i)), \tag{10}$$

by changing the parameters used in a given specification. Although the three models have significantly different specifications, using the same objective function allows us to analyse their fit by simply comparing the estimated $-\log(\mathcal{L})$ (the lower the better).

In the estimation for level 1, we assumed that the preference is crisp and equal to 0 for every dimension. When estimating the parameters of the crisp specification, we imposed the constraints that $\alpha_{i,j}$ are positive and non decreasing in j for every dimension i. When estimating the parameters of the fuzzy specifications we imposed the following constrains: (a) $l_{i,j}$ and $h_{i,j}$ are positive, (b) for every dimension i and level j, $h_{i,j} \geq l_{i,j}$, (c) the mean of $l_{i,j}$ and $h_{i,j}$ is non-decreasing in j for every dimension i. They reflect the assumptions made in the crisp model and only add the fuzzy set consistency restriction (b).

The problem specified in eq. (10), subject to the above constraints, was solved using the Nelder-Mead optimization in both cases: crisp and the new fuzzy FMN specification. The constraints were imposed by adding constraint violation penalty. In the latter, in order to ensure an effective estimation of the objective function, we replaced functions $l_{i,j}$ and $h_{i,j}$ by $(h_{i,j} + l_{i,j})/2$ and $(h_{i,j} - l_{i,j})/2$, as the latter form was simpler to impose constraints on and, in consequence, it provided a higher stability of the results. In the JKL specification, eq. (10) is non-differentiable due to the way we compare fuzzy sets. Therefore, Nelder-Mead or other standard optimization routines failed to consistently converge. To overcome these problems, we employed simulated annealing optimization in this case.

For the Nelder-Mead optimization, we used the implementation that follows Gao et al. (2012) and uses default parameters and stopping criteria, except for the number of iterations (we did not impose any restriction on the number of iterations of the optimization procedure). It was implemented in the Julia language (Bezanson et al., 2017) using the Optim package (White et al., 2017). The specification of the crisp model is effectively a logistic regression and thus it has a single local (and thus global) optimium (Menard,2002).

The FMN model, with $(h_{i,j} + l_{i,j})/2$ and $(h_{i,j} - l_{i,j})/2$ as decision variables, is a logistic regression with an additional monotonic transformation given by (9). We do not have a proof that this transformation has a single minimum, but intuitively the properties of the objective function should not be significantly different. To verify the stability of the solution, we ran the optimization 1000 times, each time starting from a new point sampled uniformly from the admissible region; the optimization converged to approximately the same solution.

For the estimation of the JKL model parameters, we found that this model has multiple local minima. Therefore, we developed a custom algorithm based on simulated annealing (Du and Swamy, 2016). The exact procedure was the following. We started with an admissible point sampled uniformly. Then, 10,000 steps of simulation annealing were performed with the application of Gaussian perturbations to all parameters (inadmissible perturbations were rejected). In the second stage, to perform a search near the optimum, we performed a random local search in which in each step we perturbed only one parameter and accepted the new solution only if it improved the solution. The second step was halted when for a batch of 1000 iterations the improvement of the objective function was less than 10^{-8} (approximately the square root of the precision of IEEE 754 floating point around 1.0). To verify that we do not end in a local minimum, we applied the multi-start (Martí, 2003) approach – the procedure was run 1000 times starting from a different random point. We report the best solution found. As with any heuristic approach, this is only an approximation of the optimal solution. However, it should be noted that better properties of the optimized objective function are another reason for preferring the FMN fuzzy approach presented in this paper over the JKL model.

4 Results

4.1 Crisp model

The results for the benchmark model are presented in Table 1, in the specification without and with the constant term α_0. In both approaches, the AD dimension was found to cause the greatest disutility (when at level 5) followed by PD; on the other hand, the UA dimension causes the least disutility (looking at level 5 only).

The estimated values of τ show that durations measured using various time units are not simply algebraically recalculated, e.g., into years. For example, $\tau_1 = 0.123 > 1/365$, showing that the relative importance of one day, when duration is measured in days, is greater than if it was measured in years.

Notice that even $\tau_2 > \tau_3$, suggesting that one week has greater weight than one month, but that may be due to estimation imprecision.

The measure of fit, that is, negative log of likelihood, has no direct interpretation (it will be compared to the ones obtained for other models).

The values presented in the table can be used to calculate the utility for all 3125 states in the EQ-5D-5L descriptive system. For example, $u(55555) = = -0.668$ (for the specification without α_0).

Table 1: Crisp model, without and with the free parameter. Measure of fit equal to 52538 and 52410, respectively. Dimensions: MO = mobility, SC = self-care, UA = usual activities, PD = pain/discomfort, AD = anxiety/depression

Dimension/level	**Description**	**No α_0**	**With α_0**
Constant		—	0.128
MO2	slight problems in walking about	0.067	0.038
MO3	moderate problems ...	0.094	0.078
MO4	severe problems ...	0.241	0.210
MO5	unable to walk about	0.318	0.290
SC2	slight problems washing or dressing	0.039	0.025
SC3	moderate problems ...	0.069	0.063
SC4	severe problems ...	0.215	0.200
SC5	unable to wash or dress myself	0.319	0.297
UA2	slight problems doing usual activities	0.111	0.065
UA3	moderate problems ...	0.135	0.094
UA4	severe problems ...	0.263	0.227
UA5	unable to do usual activities	0.263	0.227
PD2	slight pain or discomfort	0.076	0.048
PD3	moderate ...	0.122	0.093
PD4	severe ...	0.358	0.328
PD5	extreme ...	0.358	0.328
AD2	slightly anxious or depressed	0.120	0.080
AD3	moderately ...	0.221	0.196
AD4	severely ...	0.410	0.367
AD5	extremely ...	0.410	0.367
T1	day	0.123	0.129
T2	week	0.385	0.400
T3	month	0.368	0.389

4.2 JKL model

In Table 2, we present the estimation results for the JKL model, as in the original publication: Jakubczyk et al. (2017). When focusing on level 5, the PD and AD dimensions were found to cause the greatest disutility. Also, AD is associated with largest imprecision of preferences: the difference between $h_{5,5}$ and $l_{5,5}$ amounts to almost 0.6, nearly two thirds of the difference in utility between dead and full health.

In JKL specification, the utility of the worst state, $u(55555)$, is a wide interval: $[-2.02; -0.07]$.

Importantly for the present paper, the fit decreases significantly when compared with the crisp approach, even though the number of parameters has doubled. This finding motivates trying another fuzzy approach, as specified in subsection 2.3, whose results are presented subsequently.

Table 2: Fuzzy model, dimensions as in Table 1, $\rho = 0.989$. Measure of fit equals 60971

Dimension	**Parameter**	**Level**			
		2	**3**	**4**	**5**
MO	l	0.034	0.034	0.200	0.320
	h	0.215	0.215	0.500	0.601
SC	l	0.000	0.026	0.116	0.208
	h	0.186	0.278	0.388	0.530
UA	l	0.018	0.018	0.138	0.138
	h	0.138	0.206	0.355	0.389
PD	l	0.000	0.071	0.210	0.266
	h	0.296	0.296	0.546	0.771
AD	l	0.031	0.091	0.091	0.138
	h	0.120	0.242	0.701	0.727

4.3 Fuzzy model

In Table 3, we present the estimation results for the new fuzzy approach, specified in the present paper. As in the results from JKL, AD and PD are the most important dimensions (the disutility of level 5). SC and UA are associated with largest imprecision of level-5 disutility (the difference between lower and upper disutility).

Due to the approach to estimation (defining constraints on and estimating the middles and lengths of $[l, h]$ intervals, rather than l and h), the parameters l

are non-increasing in several cases (UA5, PD5, AD5). The size of this effect is small, e.g. as compared to the estimation error (not presented here).

In this specification, the utility of the worst state, $u(55555)$, is an interval: $[-1.3; -0.57]$, much narrower than in JKL.

Most importantly, the measure of fit for the newly specified fuzzy approach clearly outperforms the two earlier specifications.

Table 3: Fuzzy model, dimensions as in Table 1, $\tau_1 = 0.406$, $\tau_2 = 1.229$, and $\tau_3 = 1.265$, $\omega = 0.293$. The measure of fit is 50392

Dimension	**Parameter**	**Level**			
		2	**3**	**4**	**5**
MO	l	0.022	0.033	0.234	0.300
	h	0.125	0.157	0.309	0.433
SC	l	0.000	0.000	0.167	0.247
	h	0.046	0.111	0.235	0.441
UA	l	0.080	0.036	0.265	0.181
	h	0.124	0.216	0.296	0.380
PD	l	0.060	0.064	0.473	0.459
	h	0.123	0.252	0.503	0.517
AD	l	0.044	0.183	0.404	0.382
	h	0.134	0.229	0.506	0.528

5 Discussion

We tested the quality of three approaches for modelling the preferences towards health states. We found that the FMN fuzzy-based approach specified in the present paper clearly outperforms the other two. On the other hand, the fuzzy-based approach suggested previously by JKL performed worst in terms of model fit. The relative differences between the measures of fit are quite large, as compared to the impact of adding/removing a constant term in the crisp specification.

These results suggest the following, in our opinion. Fuzzy modelling of preferences in the context of health not only has face validity (people find it difficult to introspectively determine their own preferences) but also performs better in terms of objective criteria. We attribute those difficulties to the fact that choosing between health states forces to consider conflicting objectives: respondents have to compare (a) different dimensions of health (e.g. juxtaposing mental and physical disabilities) and (b) different durations of remaining in a given health state. Still, it is very important to introduce fuzziness properly, so as to correctly model the imprecision and its impact on decisions.

JKL's idea to model the DCE data with fuzzy sets was correct, but the concrete specification can be improved. JKL based the probability of choosing a given alternative on the difference between the membership functions calculated along the Y axis. This approach had two important features: it directly corresponded to the constant proportional trade-off (CPTO) assumption (scaling the durations should not change the preferences between the alternatives) and allowed full certainty of choice when two alternatives differ substantially. Its poor performance may be due to the violation of CPTO in empirical data (Attema et al., 2010; Jakubczyk et al., 2017), and to the failure of the exponential Bradley-Terry function used in the crisp specification to follow the CPTO.

The new fuzzy-based specification resembles the crisp approach more directly, in that the utilities of the compared health states are subtracted (the middles of 1-cuts of utilities, to be precise) to derive the probability of one state being chosen. The fuzzy approach allows the imprecision in preferences to impact the behaviour of respondents in that the probability is shifted towards 50%. The improvement in model fit (along with estimated $\omega > 0$) suggests this mechanism may be in place.

We acknowledge that many improvements can still be made to all the specifications (especially to the crisp model, which contains fewer parameters), e.g. time can be handled non-linearly. Also, the approach to the estimation process (e.g. the monotonicity constraints) could be changed (e.g. to guarantee the monotonicity in rows in Table 3). Therefore, the result of the present comparison should not be treated as the ultimate test determining the correct approach, but rather to indicate the ideas to be pursued in subsequent research.

The individual results are quite consistent between the considered approaches: pain/discomfort and anxiety/depression are the most important dimensions (i.e. the worsening to level 5). In the new specification (preferred due to predictive validity over the JKL), the SC and UA dimensions are associated with especially large imprecision (particularly when measured relative to the disutility), i.e. the difference between h and l. This may result either from the fact that 'self care' and 'usual activities' are the most vague notions in the descriptive system, or from the fact that the respondents find it most difficult to assess the value of performing these activities. It might also be the case that the importance of these criteria varies with the duration (see Jakubczyk et al., 2017). Here, we would like to highlight that the obtained results show the ability of the fuzzy approach to capture this differing uncertainty of comparing conflicting criteria in a multiple-objective setting.

Our analysis provides insight into how time – yet another dimension in the multiple-criteria decision setting considered here – is perceived when comparing health states (in the crisp approach and the new fuzzy approach). The parameters τ measure the relative importance of a unit of time (relative to 'year' as a unit). All the shorter units (days, weeks, months) in both approaches were found to have larger weight than it would follow from their actual duration. For example, in the crisp approach, τ_1 (corresponding to 'day') amounts to 0.123 and in the fuzzy approach, to 0.406, while one day equals $1/365$ of a year. There are at least two possible interpretations. First, when the problem is presented in days, the decision maker changes his/her own attitude (the framing effect) and realizes that even individual days matter. The decision maker may pay attention to the relative, not only absolute, differences in duration between the alternatives (see Jakubczyk et al., 2017).

Second, when faced with a decision problem and overwhelmed with the amount of information about conflicting multiple objectives, the decision maker may focus on numbers (how many units of time will I live in this state) and not on units (what is the actual duration). In the extreme case, if the units are neglected altogether, we would expect all parameters τ to be equal to 1.

Based on our findings, in future research it would make sense to test other approaches to time handling. For example, the duration could enter the equations in non-linear form (see also Jakubczyk et al., 2017), to drop the CPTO assumption altogether. Also, other treatments of imprecision in eq. (8) could be considered: it is not obvious that the imprecisions should depend linearly on time in the same fashion that the utility does. No estimation errors were presented in the present study. A more systematic treatment of statistical significance of findings could help to direct further research. Finally, the comparison of models was based on the model fit. Perhaps the predictive validity (out-of-sample prediction success) could be a more reliable approach.

Acknowledgments

The research was financed by the funds obtained from National Science Centre, Poland, granted following the decision number DEC-2015/19/B/HS4/01729.

References

Attema A.E., Brouwer W.B. (2010), *On the (not so) Constant Proportional Trade-off in TTO*, Quality of Life Research, 19(4), 489-497.

Bansback N., Brazier J., Tsuchiya A., Anis A. (2012), *Using a Discrete Choice Experiment to Estimate Health State Utility Values*, Journal of Health Economics, 31, 306-318.

Beresniak A., Medina-Lara A., Auray J.P., De Wever A., Praet J.C., Tarricone R., Torbica A., Dupont D., Lamure M., Duru G. (2015), *Validation of the Underlying Assump-*

tions of the Quality-adjusted Life-years Outcome: Results from the ECHOUTCOME European Project, Pharmacoeconomics, 33(1), 61-69.

Bezanson J., Edelman A., Karpinski S., Shah V.B. (2017), *Julia: A Fresh Approach to Numerical Computing*, SIAM Review, 59, 65-98.

Bleichrodt H., Wakker P., Johannesson M. (1997), *Characterizing QALYs by Risk Neutrality*, Journal of Risk and Uncertainty, 15, 107-114.

Brooks R., De Charro F. (1996), *EuroQol: The Current State of Play*, Health Policy, 37, 53-72.

Du K.-L., Swamy M.N.S. (2016), *Search and Optimization by Metaheuristics*, Springer.

Gao F., Han L. (2012), *Implementing the Nelder-Mead Simplex Algorithm with Adaptive Parameters*, Computational Optimization and Applications, 51, 259-277.

Herdman M., Gudex C., Lloyd A., Janssen M., Kind P., Parkin D., Bonsel G., Badia X. (2011), *Development and Preliminary Testing of the New Five-level Version of EQ-5D (EQ-5D-5L)*, Quality of life research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 20(10), 1727-1736.

Jakubczyk M. (2009), *Impact of Complementarity and Heterogeneity on Health Related Utility of Life*, Central European Journal of Economic Modelling and Econometrics, 1, 139-156.

Jakubczyk M. (2015), *Using a Fuzzy Approach in Multi-criteria Decision Making with Multiple Alternatives in Health Care*, Multiple Criteria Decision Making, 10, 65-81.

Jakubczyk M., Craig B.M., Barra M., Groothuis-Oudshoorn C.G., Hartman J.D., Huynh E., Ramos-Goñi J.M., Stolk E.A., Rand-Hendriksen K. (2017), *Choice Defines Value: A Predictive Modeling Competition in Health Preference Research*, Value in Health, doi: 10.1016/j.jval.2017.09.016.

Jakubczyk M., Kamiński B. (2017), *Fuzzy Approach to Decision Analysis with Multiple Criteria and Uncertainty in Health Technology Assessment*, Annals of Operations Research, 251, 301-324.

Jakubczyk M., Kamiński B., Lewandowski M. (2017), *Eliciting Fuzzy Preferences Towards Health States with Discrete Choice Experiments* [in:] Studies in Systems, Decision and Control. Vol 125. Complex Systems: Solutions and Challenges in Economics, Management and Engineering, eds. C. Berger-Vachon, A.M.G. Lafuente, J. Kacprzyk, Y. Kondratenko, C.F. Merigó, Springer, 131-146.

Martí R. (2003), *Multi-Start Methods* [in:] *Handbook of Metaheuristics*, eds. F.W. Glover, G.A. Kochenberger, Springer, 355-368.

Menard S.W. (2002), *Applied Logistic Regression (2nd ed.)*, SAGE.

Miyamoto J.M., Wakker P.P., Bleichrodt H., Peters H.J.M. (1998), *The Zero-Condition: A Simplifying Assumption in QALY Measurement and Multiattribute Utility*, Management Science, 44(6), 839-849.

Pettitt D.A., Raza S., Naughton B., Roscoe A., Ramakrishnan A., Ali A., Davies B., Dopson S., Hollander G., Smith J.A., Brindley D.A. (2016), *The Limitations of QALY: A Literature Review*, Journal of Stem Cell Research & Therapy, 6(4), 334.

Weinstein M.C., Torrance G., McGuire A. (2009), *QALYs: The Basics*, Value in Health, 12 (S1), S5-S9.

White J.M., Mogensen P.K., Holy T., Riseth A.N., Lubin M., Stocker Ch., Ortner Ch., Johnson B., Noack A., Yu Y., Carlsson K., Lin D., Covert T.R., Rock R., Regier J., Kuhn B., Williams A., Ryan, Smith D., Anantharaman R., Gomez M., Revels J., Dunning I., MacMillen D., Rackauckas Ch., Legat B., Levitt A., Stukalov A., Petrov A., Mahajan A. (2017), *JuliaNLSolvers/Optim.jl*, doi: 10.5281/zenodo.1035790.

Zadeh L. (1965), *Fuzzy Sets*, Information and Control, 8(3), 338-353.