COLLEGIUM OF ECONOMIC ANALYSIS
WORKING PAPER SERIES

(Belief in) life after death impacts the utility of life before it – a difference in preferences or an artefact?

Michał Jakubczyk, Dominik Golicki and
Maciej Niewada

Noname manuscript No.
(will be inserted by the editor)

(Belief in) Life After Death Impacts the Utility of Life Before It—a Difference in Preferences or an Artefact?

Michał Jakubczyk · Dominik Golicki · Maciej Niewada

Received: date / Accepted: date

Abstract (Purpose) In most of the religions the preservation of own, God-given, life is obligatory. The time-trade-off method (TTO) forces to voluntarily forego life years. We verify if this is a problem for the religious and how it impacts the TTO results.

(Methods) We used the data from the only EQ-5D valuation in Poland (2008, three-level, 321 respondents, 23 states each)—a very religious (mostly catholic) country. We used the *belief in afterlife* question to measure the religiosity on two levels: strong (*definitely yes*) and some (also *rather yes*), both about a third of the sample.

(Results) The religious on average (yet, not statistically significant) spend more time doing TTO and consider it more difficult. The religious more often are non-traders: odds ratio (OR)=1.97 (strongly), OR=1.55 (rather); and less often consider a state worse-than-death: OR=0.67 (strongly), OR=0.81 (rather). These associations are statistically significant ($p^* < 0.001$) and hold when controlling for possible confounders. Strong religiosity abates the utility loss: in the additive approach by 0.136, in the multiplicative approach by the factor of 2.08 (both $p* < 0.001$). Removing the effect of religiosity from the value set reduces the utility by 0.046 on average.

(Conclusions) The impact of religiosity seems to be a TTO-artefact rather than a true difference in preferences (testing this requires further analysis of, e.g., discrete-choice or visual analogue scale data). Non-Weltanschauung-biased estimates should rather be used in cost-utility analysis to drive resource allocation.

M. Jakubczyk
Decision Analysis and Support Unit, Warsaw School of Economics
Al. Niepodległości 162, 02-554 Warsaw, Poland
E-mail: michal.jakubczyk@sgh.waw.pl

D. Golicki
Department of Experimental and Clinical Pharmacology, Medical University of Warsaw
Banacha 1B, 02-097 Warsaw, Poland

M. Niewada
Department of Experimental and Clinical Pharmacology, Medical University of Warsaw
Banacha 1B, 02-097 Warsaw, Poland

Keywords Health-related quality of life · Utility · Preference elicitation · Time trade-off · Religion · Life after death

1 Introduction

Decision makers need to measure the effectiveness of health technologies, so as to decide which to finance using scarce budget. In particular the improvements in the longevity and quality of life (QoL) need to be quantified. That is often done within the quality-adjusted life years (QALY) model (Bleichrodt et al 1997): the health states are assigned numerical values, interpreted as utilities of spending a year in that health state (the perfect health state is normalized to 1, and death to 0).

Assigning utilities requires making the 'health state' concept more operational. Often the EQ-5D-3L approach is used: each health state is described using five dimensions (mobility—MO, usual activities—UA, self-care—SC, pain/discomfort—PD, and anxiety/depression—PD), each evaluated on three levels, 1–3 (Brooks and De Charro 1996).[1] A given health state, e.g., 21121, is then assigned utility, often using a time trade-off (TTO) method: a thought experiment in which the respondent compares living 10 years in 21121 versus $t \leq 10$ years in full health (11111 in EQ-5D-3L parlance), both followed by immediate death. Various t are tried, until t^* is found for which the respondent is indifferent between the two profiles. We then assign to 21121 the utility of $\frac{t^*}{10}$. The results of TTO are usually extrapolated from a smaller subset to all the 243 (3^5) health states via econometric modelling (Dolan et al 1996).

Importantly for the present paper, TTO forces the respondent to voluntarily forego some life years (by allowing reductions in t in an iterative process described above) in exchange for QoL. That hastening one's death may be, subconsciously perhaps, perceived as a form of a suicide—mentally an uneasy task. According to major religions suicide is a grave sin. Christianity, the most popular religion in Europe, states in the Sixth Commandment: *"Thou shalt not kill"* (Exodus 20:13), which also forbids killing oneself. The preservation of own life is obligatory as it is God-given: *"... The Lord gave, and the Lord has taken away..."* (Job 1:21). Dante in "Divine Comedy" reserved the seventh circle in Hell for the suicides to be transformed into bushes and fed upon by Harpies. Quran, the holy book of Islam (the second most popular religion in Europe), says: *"Nor kill nor destroy yourselves, for surely God has been to you Most Merciful"* (An-Nisaa' 4:29). Back to a more empirically-based discussion: Danyliv and O'Neill (2015) using the British Social Attitudes survey collected in 1983–2012 showed that the more religious (measured as church attendance) oppose euthanasia more, and religiosity is the single most important predictor of this attitude. Borrill et al (2011) in a sample of British students discovered an association between religiosity and repeated self-harm.

The goal of the present study is to i) empirically verify whether religiosity impacts the results of the TTO experiment on the individual respondent level, ii) measure the strength of this impact, also for the value sets, and iii) discuss the consequences for the utility elicitation methods. If the religious on average report higher t^* values than non-believers, then this results in higher utility values. This

[1] And for several years EQ-5D-5L has been being developed (Herdman et al 2011).

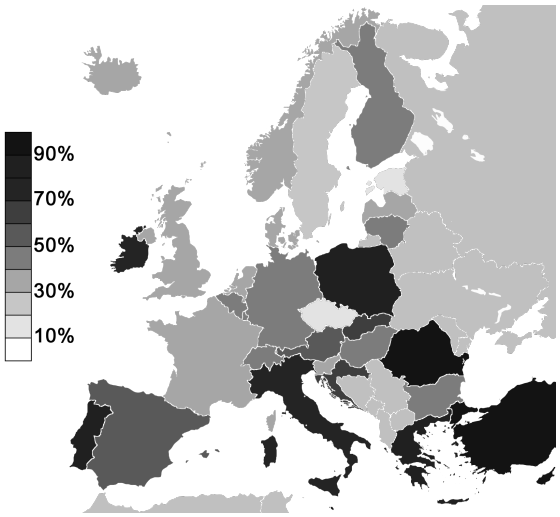

Fig. 1: Percentage believing in God's existence in the European Union (licensed under CC BY-SA 3.0 via Wikimedia Commons—`http://commons.wikimedia.org/wiki/File:Europe_belief_in_god.svg`).

may be perceived as an artefact of TTO as a utility elicitation method rather than the actual difference in preferences. Thus, understanding the phenomenon is important for determining the proper target group for elicitation tasks, selecting the proper elicitation method, designing the protocol (e.g., asking about religiosity), or selecting proper methods to analyse the data (e.g., controlling for religiosity in modelling).

We use data coming from a Polish EQ-5D-3L valuation study (Golicki et al 2010). As much as 93.1% Poles declare to be Catholics (Boguszewski 2012), 56% of Poles do not have any doubt there is God (27% have some doubts and only 3% declare an absolute disbelief), and 67% believe in life after death (36% in the heaven-hell-limbo version, 31% in an unspecified fate) (Boguszewski 2015). Poland stands out from the other European Union (EU) countries in terms of religiosity measured as belief in God, as shown (also for some non-EU countries) in Fig. 1 (Eurobarometer 2005). Poland is also prominent with respect to the active religiosity: 63% Poles weekly attend religious services, as compared to 33% in Slovakia, 14% in Lithuania, 11% in the Czech Republic, and below 10% in Germany (Gallup 2004). Hence, using Polish data seems a natural choice. In the valuation study the respondents were asked about the belief in life after death (henceforth BLD) rather than simply 'being religious', which can be understood as a proxy of a more conscious religiosity. Jakubczyk (2009) analysed the association between the demography and how the health states are valued, and showed that BLD is associated with smaller reported disutility of worsening the MO, UA, and PD dimensions, yet the issue was not pursued in detail.

2 Methods

We used data coming from the only EQ-5D-3L valuation study in Poland using TTO. Golicki et al (2010) present the details of the study. Each of 321 respondents (visitors of inpatients) self-reported own health state using EQ-5D-3L, performed the ranking exercise, valued the ranked states using VAS, rated own state with VAS, valued 23 states using TTO, and answered general socio-economic questions. In total 44 states were used, split into two subsets of 23 states. Using numerous health states per respondent was shown not to be a problem (Golicki et al 2010, 2013).

In the demographic section the respondents were asked about BLD. We trust that this reflects better the actual attitude towards religion in Poland: asking for *being a Catholic* or *believing in God* might give many positive, a bit automatic, answers. Asking for BLD is directed at a more thoughtful religiosity. The possible answers encompassed *definitely yes*, *rather yes*, *rather no*, *definitely no*, *no answer*, and *I don't know*. Using this question in the present study we measure the religiosity on two levels: strong (*definitely yes*) and some (also *rather yes*). We treated *no answer* as missing data and *I don't know* as a negation (*I don't know* is considerably weaker than *rather yes*).

We first verified, in the complete data set, how religiosity (strong, some, none) is associated with demographic characteristics: sex, age, place of residence (city vs countryside; Warsaw vs not), education (higher vs other), having a partner/spouse (vs being divorced, a widow, etc.), occupation (working vs other), and income (greater than 2,000 PLN/month or not). That served to decide what demographic variables should be controlled for subsequently, so as to reduce the risk of confounding. We used χ^2 test for binary variables and ANOVA for continuous ones. We also checked how religiosity is associated with the duration of the interview, the perceived (by the pollster) difficulty with ranking/VAS/TTO (1—very easy, to 5—very difficult), and the self-reported VAS (using ANOVA for duration and VAS, and Kruskal-Wallis for difficulty).

The data set was then restricted, similarly to Golicki et al (2010): we removed respondents with many serious inconsistencies (≥ 10 pairs of valuations violating Pareto-dominance by more than 0.5) and then removed individual valuations differing by more than 3 standard deviations (SD) from the mean valuations. Golicki et al (2010) used 2 SD, yet in the current study we decided to keep more varying answers as detecting variability between answers was the main goal of the study. All the remaining analyses were done in the restricted data set.

We started with comparing the frequency of considering a state worse than 11111 (i.e., being a trader) and the frequency of considering a state WTD (both using a χ^2 test) the religious (separately on two levels: strong and some) and non-religious. As religiosity was not a stratification factor, so as to control for the impact of states being valued, we performed a stratified Cochran-Mantel-Haenszel (CMH) test. In order to control for other demographic variables, we used a logistic regression using as independent variables the religiosity, dummy variables for health state being valued, and (selected based on initial analysis) demographic variables. As our goal was not to construct the actual model, we did not remove variables that turn out as insignificant, but focused on the impact of religiosity.

The remaining analyses were performed for strongly religious vs the other (this threshold was set based on the analyses described above). We compared average

valuations and variations between the religious and not, separately for individual states (t-test, unequal variances, and F-test). We used Holm-Bonferroni correction to control the family-wise risk of type I error in this multiple hypotheses testing. In order to quantify the average impact of religiosity, we built linear models explaining the utility loss with state being valued (dummy variables), religiosity, and demographic variables. We used an additive model (religiosity impacts the valuation in absolute way) and a multiplicative model (religiosity impacts the utility loss attributed to a given health state as compared with 11111 in relative terms) using semi-logarithmic model. We used random-effects panel models to account for intra-respondent correlation in error terms (we had to remove 8 valuations as respondent-state pair was mistakenly repeated). In the multiplicative model we replaced zero loses with 0.001 (half of the smallest non-zero loss), to avoid problems with taking the log. We used the Box-Cox transformation to see whether multiplicative or additive model fits the data better.

In order to compare the impact of religiosity also on the values sets obtained via extrapolation, we built and compared four linear models (panel modelling, random effects): M1) for a complete set, treated as a benchmark, M2) with the strong-believers removed, M3) for all the respondents with known religiosity, with dummy variable for religiosity, and M4) for all the respondents with known religiosity, with dummy variable for religiosity and additional interaction variable between the religiosity and the misery index (decreased by 5) to allow for changing reduction depending on the gravity of the health state. We treat M4 as the recommended specification that allows to remove the impact of religiosity. We extrapolated the results of M1–M4 to all the 243 health states (for a non-religious person) and visually compared the results.

In statistical testing we generally used two-tail approach (where possible) and assumed $\alpha = 0.05$ as a significance level (we report exact p^*-values except when < 0.001). The calculations were done in Statistica 10 and R 3.1.0.

3 Results

Among 321 respondents 2.5% did not provide an answer to the belief in afterlife question (and were removed, except for when building the M1 model); 9.66% did not know (and were assumed as non-believers). Ultimately, in our sample 29.7% definitely believed, 34.5% rather believed, and 35.8% (rather or definitely) did not believe in afterlife. Due to this almost exactly one-third split we decided to start with analysing the impact of both—a strong and some belief in afterlife, so as to select where the appropriate threshold lies.

Table 1 shows how religiosity is associated with other demographic characteristics: it is significantly more prevalent among women, living in a countryside, not working, and having a low income. The strength of association consistently increases with the intensity of religiosity. The sample data encompassed many respondents from Warsaw but, luckily, that is not associated with religiosity. Surprisingly, religiosity is associated with the working status, probably due to both being correlated with other variables, e.g., age. For further discussion notice that, however non-significant, there were some trends in how the religious found the comparing tasks (ranking and TTO) more difficult, also making the overall survey more time consuming.

Table 1: Religion, demography, and difficulty (greater values denote greater difficulty). Significant p-values marked with an asterisk.

variable	definitely yes	rather yes	no	whole data set	p^*
Binary variables: % of respondents, χ^2 test p-value					
subset 1 of valued states	57.0%	57.0%	47.3%	53.0%	0.259
male	38.7%	44.4%	57.1%	47.0%	0.024
countryside	20.4%	16.7%	8.0%	14.6%	0.034*
Warsaw	58.1%	59.3%	68.8%	61.7%	0.210
spouse/partner	77.4%	64.8%	72.3%	71.7%	0.137
higher education	39.8%	38.0%	46.4%	42.4%	0.409
working	51.1%	70.4%	65.2%	62.9%	0.016*
high income	27.8%	35.9%	46.2%	36.1%	0.029*
Continuous variables: mean (SD), ANOVA or Kruskal-Wallis p-value					
age [years]	46.1 (16.6)	41.3 (14.9)	41.5 (15.4)	42.8 (15.7)	0.054
duration [minutes]	44 (15.9)	46.5 (39.1)	39.6 (12.3)	43.3 (25.9)	0.151
difficulty ranking	2.28 (0.92)	2.29 (0.83)	2.11 (0.85)	2.22 (0.87)	0.442
difficulty VAS	2.02 (0.85)	2.02 (0.83)	1.99 (0.94)	2.01 (0.87)	0.693
difficulty TTO	2.48 (1.06)	2.43 (1.01)	2.27 (1.06)	2.39 (1.05)	0.384
self-reported VAS	80.3 (15.2)	82.3 (12.5)	81.2 (15.5)	81.3 (14.4)	0.599

TTO: time trade-off, VAS: visual analogue scale,
SD: standard deviation, ANOVA: analysis of variance

The religious more often are non-traders (unwilling to give up any time in TTO) and so effectively treat the valued state as perfect health (assign utility equal to 1). The strong believers didn't trade in 21.03% valuations, some believers in 12.48% cases, and non-believers in 11.39% cases. Odds ratio (OR) amounts to 1.97 for strong believers (vs others) and OR=1.55 for (at least) slight believers (vs others), for both comparisons χ^2 $p^* < 0.001$. CMH test stratified for health state being valued confirmed this: OR=2.07 and OR=1.59, respectively (both $p^* < 0.001$). Full logistic model including also the dummy variables for health state, sex, age, living in countryside, working, and having a high income further confirmed the results yielding OR=1.9 and OR=1.37, respectively (both $p^* < 0.001$).

The religious less willingly consider a state WTD (effectively, assign utility < 0): the strong believers did so in 19.05% valuations, some believers in 25.57% cases, and non-believers in 26.26% cases. Then, OR=0.67 (strong, $p^* < 0.001$) and OR=0.81 (some, $p^* < 0.001$). Again, CMH test stratified for health state being valued confirmed this: OR=0.58 and OR=0.72, respectively (both $p^* < 0.001$). Full logistic model (analogous to the previous one) yielded OR=0.57 and OR=0.78, respectively (both $p^* < 0.001$). It can be seen that the strength of the belief impacts the results, but the threshold seems to lie between the strong and some believers. Therefore the subsequent analyses are performed for strong believers vs the others (some or non-believers combined).

We compared the mean valuations of individual states between the strong believers and the others (Table 2). The difference is statistically significant for almost half of the states when each hypothesis is treated individually. Due to multiple-hypothesis testing we introduced Holm-Bonferroni correction, and the difference was found significant in five cases (and so the family-wise conclusion is

that religiosity impacts the health state valuation). Notice that in every single case the average utility was higher for the religious.

In Table 4 in the Appendix we present additionally the comparison of the mean valuations in cases where some time was traded-off in TTO. Using Holm-Bonferroni correction we do not reject the null hypothesis that the mean is not influenced by the religion ($0.0013 > \frac{0.05}{44}$). It may be not surprising: once you only use valuations in which the respondents manage to accept the logics and rules of the thought experiment, the results do not differ between the religious and the others. In a sense the religiosity impacts the mean valuations mostly through impacting whether or not the respondent will be a trader in the first place. (In the Appendix in Table 4 we present additionally the result of the comparison of variances to illustrate some points raised in the discussion.)

Table 2: Valuations (sorted by p-value). An asterisk by the p-value denotes a standard statistical significance ($p^* < 0.05$), a double asterisk denotes significance also with Holm-Bonferroni correction.

state	the others vs strong believers	p^*	state	the others vs strong believers	p^*
33333	-0.472 vs -0.168	0.0001**	21222	0.766 vs 0.844	0.0727
32331	-0.216 vs 0.187	0.0002**	22121	0.812 vs 0.875	0.0844
32333	-0.407 vs -0.039	0.0006**	12223	0.535 vs 0.664	0.0927
23333	-0.326 vs 0.036	0.0009**	22323	0.270 vs 0.431	0.0946
22331	-0.008 vs 0.336	0.0011**	11312	0.712 vs 0.797	0.1017
33323	-0.235 vs 0.107	0.0049*	33212	0.260 vs 0.443	0.1108
32211	0.433 vs 0.684	0.0069*	12211	0.844 vs 0.892	0.1590
21312	0.534 vs 0.712	0.0073*	11211	0.914 vs 0.941	0.1660
32232	-0.112 vs 0.177	0.0086*	21111	0.919 vs 0.942	0.1897
21133	0.157 vs 0.417	0.0098*	21232	0.271 vs 0.392	0.2440
33321	-0.021 vs 0.300	0.0129*	13212	0.656 vs 0.725	0.2492
23321	0.245 vs 0.478	0.0141*	22122	0.761 vs 0.804	0.2982
32223	0.153 vs 0.409	0.0190*	11113	0.695 vs 0.752	0.3047
13332	-0.121 vs 0.134	0.0217*	12111	0.906 vs 0.930	0.3888
22112	0.776 vs 0.857	0.0322*	22233	0.049 vs 0.152	0.3956
22222	0.694 vs 0.771	0.0351*	23232	0.040 vs 0.132	0.4287
11133	0.167 vs 0.376	0.0440*	13311	0.498 vs 0.554	0.5566
21323	0.414 vs 0.595	0.0499*	12121	0.864 vs 0.881	0.6338
23313	0.108 vs 0.339	0.0502	11112	0.909 vs 0.919	0.6814
32313	-0.005 vs 0.210	0.0530	11122	0.826 vs 0.844	0.7158
33232	-0.230 vs 0.008	0.0533	11121	0.899 vs 0.906	0.8244
12222	0.742 vs 0.818	0.0673	11131	0.322 vs 0.329	0.9501

In order to have a joint measure of the impact of religiosity on valuations, we built two regression models. Strong religiosity abates the utility loss: in the additive approach on average by 0.136 and in the multiplicative approach by the factor of 2.08 (both $p^* < 0.001$, also when heteroskedasticity-consistent estimation method is used). Using non-panel, ordinary least square regression with no observations removed changed the results only a little. Box-Cox method suggested the log-likelihood maximizing value of $\lambda \approx \frac{1}{3}$, and so the multiplicative approach should perhaps be slightly favoured. That is in line with the results presented in Table 2, where the absolute difference seems to be greater for worse health states.

Finally, we checked the impact of religiousness on the value sets. To do that we build four linear models to extrapolate the valuations onto all the health states. In spite of the above results, we used an additive approach, being a standard one in such tasks. The coefficients measuring the impact of individual domains are greater when the influence of religion is removed, cf. Table 3. The results of M1–M4 extrapolation are illustrated in Fig. 2. The mean utility is lower in M4 by 0.046 than in M1 (treated as a benchmark).

Table 3: Coefficients of the models M1–M2 (SF—strong faith, MI—misery index).

variable	M1	M2	M3	M4
Intercept	0.068674	0.071905	0.117205	0.0568098
MO2	0.056190	0.063230	0.053062	0.0638144
MO3	0.313355	0.347513	0.308310	0.3306477
SC2	0.049534	0.060317	0.053207	0.0667350
SC3	0.226134	0.238243	0.227218	0.2511037
UA2	0.063239	0.068394	0.059850	0.0719717
UA3	0.223958	0.253568	0.223056	0.2484439
PD2	0.051498	0.059646	0.051187	0.0632948
PD3	0.463747	0.492456	0.463710	0.4888381
AD2	0.022164	0.016572	0.025069	0.0364197
AD3	0.200485	0.202717	0.198738	0.2219464
SF	—	—	-0.162655	0.0306256
SF × (MI−5)	—	—	—	-0.0384869

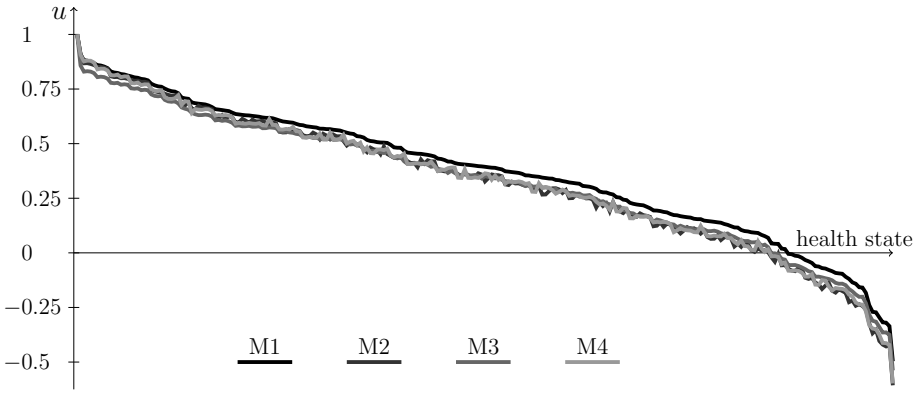

Fig. 2: Value sets build for M1–M4 (health states ordered by values of M1).

4 Discussion

Our results show that the religious (at least Catholics in Poland) assign greater utilities to health states when surveyed using TTO than non-religious. Our interpretation is that this is only an artefact of the TTO methodology. TTO runs as

an iterative process, and considering a full health profile being better results in it being shortened in next round. In this sense conducting TTO is asking to forego some life years, and the religious are more reluctant to this (and more often are complete non-traders). Thus the utility assigned to the valued health state is effectively increased. Another explanation could, however, be given. It might be that the religious simply do attach (on average) greater utility to all the health states (in other words—perceive the utility loss resulting from worsening the QoL as smaller). That might stem from, e.g., them finding comfort in faith when enduring worsened QoL. The latter interpretation means that QoL has smaller impact on utility (for the religious as compared to the non-religious), and it's the longevity of the health profile that counts.

Even though the two above interpretations are intuitively different, it is more difficult to define the difference between them formally on the grounds of decision theory. In economics, at least since von Neumann and Morgenstern (1944), utility is simply the representation of choices (and not, e.g., some actual stream of pleasure experienced in any physiological way). If the religious tend to select some options more frequently in standard TTO, it *means* that they *do* attach greater utilities to them. To give a meaning to the difference between the two interpretations we have to refer to other methods of utility elicitation, e.g., visual analogue scale (VAS), standard gamble (SG), and discrete choice experiment (DCE).

With VAS the respondent is asked to rate a given state by selecting a point on a thermometer-like line. The respondent is also asked to rate the immediate death 'state', and we assume that the perfect health is at the top of the thermometer. Assuming that the attractiveness of the state changes at constant pace along the thermometer scale, we can assign numerical value to each state. We call this value utility, remembering it actually comes from a sorting-type exercise and so formally can only be treated as an ordinal utility. Still, the ease of this task may compensate for the abuse of interpretation. Using VAS does not seem to involve any ethical difficulties for the religious as longevity of the profile is not considered at all. The results of a study conducted in Norway focused on attitudes towards euthanasia (details of the study have been described by Augestad et al 2013) showed that also religiosity (measured along 5 point scale) was statistically significantly associated with TTO results but not with VAS results,[2] which confirms the hypothesis that our findings are a TTO-artefact at least to some extent. As VAS is known to produce different results than TTO (Robinson et al 1997), it would require a dedicated analytic approach, however, to measure what part of our results is driven by the actual difference in preferences.

With SG a respondent selects p such that continuing to live in a given state is equivalent to taking a pill which will remove all the QoL problems (without prolonging the life, however) with probability $1 - p$, but may result in an immediate death with probability p. Then the utility of health state under valuation is given simply by $1 - p$ (Green et al 2000). With this approach we can foresee a similar ethical problem for the believers as in TTO. A religious respondent may feel reluctant to accept the gamble at all, feeling it improper to voluntarily put one's life on stake. Under our interpretation, p will be biased downwards (and more often equal to 0) for the religious, and in the result, again—greater utilities will

[2] Personal e-mail correspondence with K. Rand-Hendriksen, University of Oslo, a co-author of the Augestad et al (2013) study.

be attached to the health states. The overall picture will be blurred, however, if the religious have different risk-attitude than the non-religious (e.g., greater risk aversion would also result in p being decreased for the believers)[3].

In DCE the respondent makes a series of choices, each from (usually) two health profiles (Bansback et al 2012). Assuming some underlying utility function drives the choices, we can derive utilities for health states via econometric modelling. Making such choices is often difficult (comparing health profiles substantially differing in attractiveness would yield no valuable information); still (at least to us) whether there is a similar ethical problem for the religious in DCE remains unclear. On one hand, when profiles differing in duration are presented, the decision maker may feel it improper to select a shorter-lasting profile. On the other hand, the respondent is forced to make a choice anyway, and the profiles are not modified in the process (i.e., a duration of the more attractive profile is not reduced, contrary to TTO). Whether or not there are differences between outcomes of DCE for the religious and not (and differences in differences between DCE and TTO) should then be analysed in the future. It may get more difficult when being indifferent is allowed as an outcome. Then the religious may react by not wanting to step into God's shoes and using this answer more frequently. The impact of indifference on the final estimates (of health states utilities or impact of EQ-5D dimensions on utility) would depend on what profiles (durations and health states) are compared (e.g., comparing many profiles with similar durations could result in flattening the results, i.e., all the states/dimensions being valued similarly).

We could also try to modify TTO to verify which explanation holds. Say, we ask the respondent to compare 10 years in perfect health versus a combined profile of t years in a given health state (e.g., 21121) followed by *another* t years of perfect health. Focusing on states better than dead, for $t = 10$ obviously the mixed profile is better, and we can then reduce t so as to obtain indifference (we should obtain indifference for $t > 5$). If the decision maker is indifferent between 10 years in 11111, and t^* years in 21121 followed by t^* in 11111, then $u(21121) = \frac{10-t^*}{t^*}$. This is illustrated in Fig. 3. Differently from a regular TTO, the respondent is now asked (implicitly, again via an iterative process) to shorten the duration of the profile involving the tested health state. Should the religious oppose to the very action of shortening life, that would result in biasing t^* upwards and so reducing the final elicited utility. If, on the other hand, the religious differ in preferences, caring about the longevity of life more, they would still select lower t^* than non-believers.

For the sake of the discussion that follows, we continue with our interpretation that the religious dislike the shortening of life, and higher valuations are a TTO-artefact. This is additionally, yet very slightly, supported by Table 1: the religiosity seems not to impact in the slightest way the difficulty of the VAS task (where states are not compared to each other, but valued individually), while there is a vague hint that religiosity affects the difficulty in the ranking and TTO exercise (where states are being compared). The religious might also need more time to answer all the questions, but again—that is far from statistical significance. Another, a bit qualitative, argument is that even if the believers find support in their faith to stand health problems better[4], the EQ-5D descriptors are not defined objectively,

[3] We are being a bit informal here, as no expected value can be calculated in the space of health states, but the intuition remains clear.

[4] As suggested, e.g., by Saffari et al (2013), who showed that religiosity/spirituality improves the QoL of patients undergoing haemodialysis. However, Nagpal et al (2015) found no such

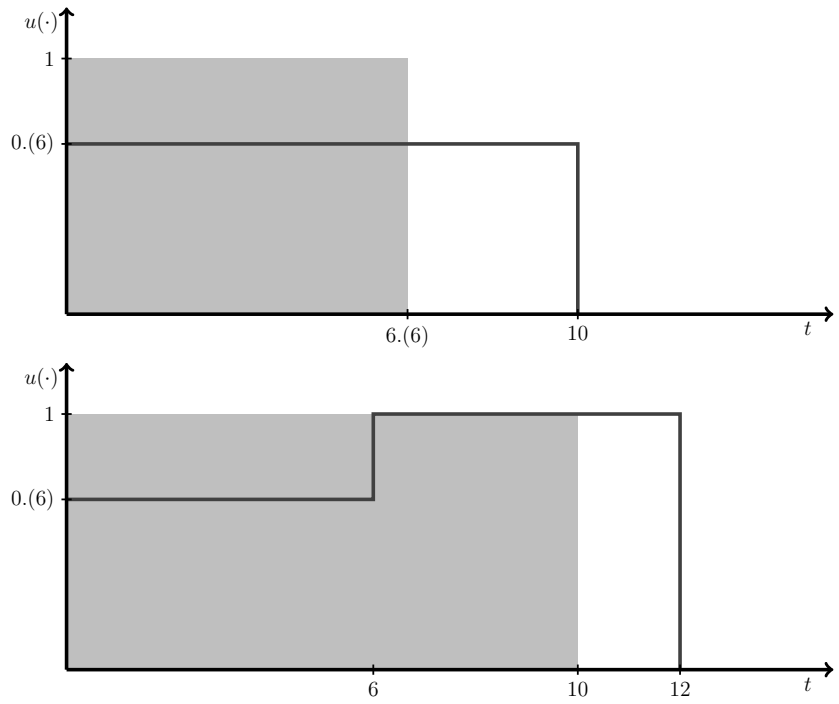

Fig. 3: A regular TTO (above): life in 11111 (grey rectangle) is lengthened/shortened so as the grey & bordered areas equal. A modified TTO (below): a tested state is a part of a composite profile (bordered polygon), which is the one to lengthen/shorten.

and so are perceived subjectively, thus the believers might simply attach a worse clinical image to the same EQ-5D description compensating for their more positive outlook on the QoL.

The two interpretations differ importantly in implications. If the believers truly attached (on average) higher utilities to health states (i.e., they would do it irrespectively of the elicitation method used), then when building value sets we would need only to think about the religiosity as one of the demographic variables that should be balance to make the sample representative. However, if it's the TTO that causes the difference, then we need to think *whose* valuations we should use in decision making (the believers', the non-believers, the average?). We take the following standpoint. The valuations are not made only out of curiosity, but to inform decisions on making health technologies available to the patients. The decisions are rather about whom to treat than about whether to shorten one's life in order to improve that persons's QoL. Hence, TTO is a very indirect way of obtaining the utility weights to be subsequently used in decisions: there is a discrepancy between what we use the data for and what kind of a thought experiment we use to get them. Hence, if TTO influences the results (by changing the answers of the

association in individuals with dementia, and Jelsma and Ferguson (2004) found no relation in general population.

believers), we should try to remove this effect from the final value sets. The utility elicitation method that is much closer to how the elicited values are then used is the person trade-off (PTO), in which benefits in various group of patients are compared; PTO is, however, difficult in practice, e.g., due to fragility to framing (Nord 1992, 1995), and not popular in practice (focusing on health gains might also make the result depend on the reference point, see Nord et al 2009).

Several methods might be used to remove the TTO-religion artefact. Most drastically, the religious could be removed from the sample altogether, which seems difficult and expensive to undertake (we do it in M2). A better, and politically correct, solution is to use some other elicitation method, and DCE looks (from that perspective) as a somewhat promising alternative. Obviously, there are other and more important criteria to be considered when selecting between, e.g., TTO and DCE, than the impact of the religion. Therefore, the third solution might be most reasonable: to collect information on the religiosity[5] and to control for the effect of this attitude on the resulting value sets via econometric modelling (e.g, as we do in Table 3 in M3&M4). The resulting value sets can differ quite substantially and will assign lower utilities to health states when the impact of religion is removed. This would have implications for the results of cost-utility analysis. The differences between health states will be generally made bigger (especially between full health and a given state), which will increase the gains from the QoL improving technologies. On the other hand the prolongation of life will be valued less. The overall average effect is ambiguous, but the decisions will be distorted towards QoL-improving technologies. Thus, however exotic the present paper might look, the implications can be quite serious. Nonetheless, just to complete the picture, let's be a bit frivolous only in the following single paragraph.

One could say that when comparing two health profiles the believers should also include the stream of utility derived from the period *after death*. To make it mathematically tractable we assume that, i) this stream is constant, ii) the momentarily utility is finite, and iii) the discounted quality-adjusted life years are being compared. Assuming exponential discounting with some coefficient ρ,[6] we conveniently obtain that the eternity amounts simply to ρ^{-1} discounted years. Then the total eternal stream of utility sums to a finite value, and we can still compare health profiles. Assuming utility of after-death is greater than 1, any $t \leq 10$ in 11111 is better than 10 years in any state, and the TTO protocol would require shortening t (trying in a counter-productive way to reduce the attractiveness of this profile) and would result in all the states being considered WTD, which is true for after-death utility being greater than 1. This is contrary to what we observe, i.e., the believers less frequently reporting a state WTD. Most likely, the after-death stream of utility is disregarded, e.g., as being something qualitatively different and so mentally accounted separately, or as being infinite after all, and removed from the picture.

More seriously, the literature also points to the difference between *being dead* (a null state in which one does not enjoy any actual health condition) and *dying* (an ultimate experience, we would rather not go through). It was tested whether the religious and not differ with respect to the fear of the latter, but the re-

[5] We only used a crude definition of religiosity, for other possibilities see, e.g., Gielen et al (2009).

[6] Utility u obtained in time t in the future is worth at the present moment $ue^{-\rho t}$.

sults are strikingly inconclusive: religiosity and fear of death being associated positively/negatively/in an inverse U-shape/not-at-all, depending on the study (Ellis et al 2013). Luckily, these inconclusive result do not have much impact on the present study as the fear of death do not seem to explain our results. First of all, death ultimately comes in both compared profiles in TTO and so should cancel out, not having an effect of results. It would not cancel out completely, however, if we account for discounting. Then the shorter profile (in full health) leads to a death coming sooner, and so a more reduced utility (assuming positive discount rate). This would result in greater elicited t^* in TTO (to compensate for a fear of a sooner death). If the religious fear death less (as usually hypothesised in this line of research), then for them this effect should be smaller and so smaller elicited t^* should follow, contrary to what we find in the data.

The religious and others may also differently perceive implausibility of the compared health profiles. The non-believers might (subconsciously perhaps) think that it is simply unrealistic to consider living in a very bad state for 10 years, as euthanasia is always an option. That might on one had lead to not-so-bad evaluation of this bad profiles by the non-believers; again, exactly opposite to what we observe in the data. We tried to see whether there are some hints of differences in the plausibility of health states (as perceived by the religious and not) measured by the variance of valuation results. Table 4 in the Appendix presents the results, which are slightly difficult to interpret. On one hand the null hypothesis that the variances differ should be rejected (also when controlling for family-wise type error), but that is only due to a difference between how 33333 health state was valued. It seems that believers differed more in their evaluations, and that may result from the negative utilities being (a bit artificially, for technical purposes only) forced into $(-1, 0)$ interval by a distance-reducing transformation (as was typically done for a regular TTO for WTD, for technicalities see Golicki et al 2013). Using the lead-time TTO (see Attema et al 2013; Devlin et al 2011), where negative utilities are not transformed and maintain interpretation, might change this outcome. Also notice that for various states the variance is either greater or lower for the religious. Thus, it is difficult to convincingly interpret the results of the variance comparison, but still, there is no strong evidence that religiosity impacts the (perceived) plausibility of health profiles.

We could also ask if our results can be driven by some third variable (e.g., some demographic characteristic) being correlated with religiosity and different perception of QoL. E.g., the elderly are religious more often and might also be more focused on the longevity rather than on the QoL. That would mean that our interpretation of the mechanism is wrong, and the conclusion would be that other elicitation methods should consistently show the difference between the valuations of believers and non-believers. In our study we controlled for demographic characteristics most commonly associated with religiosity, and the effect of religiosity remained unchanged (for further information what demographic variables are associated with so called orthodoxy index see, e.g., Boguszewski (2015)). Poland is quite a homogeneous country with respect to ethnicity, which was shown to be associated with attitude towards euthanasia (Al-Sharifi et al 2015). Poland is also very homogeneous when it comes to religion, i.e., believers are to a large extent Catholics. Therefore the results are rather not driven by any heterogeneity with respect to what kind of life after death the respondents would expect.

It was shown in the literature that demographic characteristics different than the ones in our data set can impact the results of the TTO experiment. It was found that providing care for others (children or adults) results in respondents not willing to select shorter-duration profiles (Matza et al 2014). If this is positively associated with being religious, this may pose a confounding problem. The actual causal path would then look the following: *being a believer* ⇒ *readiness to care for others* ⇒ *rejecting short profiles* ⇒ *higher TTO valuations*. That is impossible to tell based on our data set, however, there was no correlation between the religiosity and marital status (probably correlated with providing care), suggesting the above link might not explain our results. Even if this explanation holds, still the effect of religion/care giving on valuations should be removed from the value sets used to evaluate health technologies.

The sample we used was representative of the overall population in Poland only in terms of age and sex (while, e.g., numerous respondents came from Warsaw). We believe, that even though it might impact the quantitative assessment of the strength of religion, it is not likely to entirely explain the phenomenon. Another issue arises here—it was shown that medical doctors more often oppose euthanasia independently of religiosity (Seale 2009). That may suggest that being actively engaged in the process of life saving makes people value life. Hospital visitors that were surveyed may be treated as more-than-averagely conscious of the life-saving issues, which could result in upward bias of valuations irrespectively of the religion.

5 Conclusions

In the last words—it is a fact that the religious assign in TTO greater utilities to health states than non-religious. We tend to believe that this is the result of the way the elicitation is done and of the religious opposing voluntarily foregoing life-years. Then, the difference is a TTO-artefact (i.e., would disappear in some other elicitation methods). This artefact should be removed as value sets should support decisions substantially different than TTO-like thought experiment. However, under another interpretation the religious indeed value in life longevity more than quality; then the difference should be reflected in the value sets and in the decisions. Further research for data obtained with other elicitation methods is needed. Obviously, this paper absolutely should not be treated as a criticism or approval of faith in God/afterlife, suicide, or euthanasia.

Acknowledgements

This paper was partially prepared when M. Jakubczyk was visiting at The University of Iowa, thanks to the Fulbright Senior Award. The paper has benefited from the comments provided during the EuroQol Group Annual Meeting in Kraków, Poland, after the discussion initiated by H. Bailey. The views expressed by the authors do not necessarily reflect the views of the EuroQol Group.

Appendix

Table 4: Comparison of means (traders only) and variations (all) between strong believers and the others, * denotes standard statistical significance ($p^* < 0.05$), ** denotes statistical significance also with Holm-Bonferroni correction.

state	comparison of means (traders only) the others vs strong believers	p^*	comparison of variances the others / strong believers	p^*
11112	0.867 vs 0.863	0.9169	1.21	0.4492
11113	0.636 vs 0.664	0.6587	0.95	0.8166
11121	0.861 vs 0.829	0.5263	0.64	0.0865
11122	0.768 vs 0.763	0.9437	0.69	0.1062
11131	0.261 vs 0.275	0.9094	0.77	0.2976
11133	0.118 vs 0.213	0.3790	1.15	0.5944
11211	0.874 vs 0.892	0.5437	1.08	0.7663
11312	0.661 vs 0.707	0.4613	0.81	0.4197
12111	0.863 vs 0.874	0.8127	0.66	0.1107
12121	0.817 vs 0.830	0.7733	0.67	0.1185
12211	0.795 vs 0.846	0.2278	0.94	0.7807
12222	0.677 vs 0.746	0.1529	1.28	0.3337
12223	0.474 vs 0.593	0.1563	1.11	0.7352
13212	0.597 vs 0.654	0.3906	0.91	0.7017
13311	0.441 vs 0.424	0.8710	0.94	0.7861
13332	-0.176 vs 0.021	0.0709	0.96	0.8560
21111	0.881 vs 0.907	0.2935	1.03	0.9271
21133	0.098 vs 0.314	0.0389*	1.21	0.4543
21222	0.713 vs 0.777	0.2251	0.72	0.1971
21232	0.189 vs 0.268	0.4706	1.02	0.9432
21312	0.463 vs 0.628	0.0266*	1.47	0.1247
21323	0.354 vs 0.478	0.2234	1.28	0.3928
22112	0.714 vs 0.799	0.0636	1.34	0.2469
22121	0.763 vs 0.821	0.1943	0.98	0.9079
22122	0.701 vs 0.739	0.4308	1.00	0.9632
22222	0.636 vs 0.679	0.3029	1.09	0.6334
22233	0.008 vs 0.108	0.4151	0.86	0.5325
22323	0.188 vs 0.302	0.2594	1.01	0.9759
22331	-0.057 vs 0.235	0.0063*	1.22	0.4193
23232	-0.001 vs 0.087	0.4478	1.06	0.8577
23313	0.090 vs 0.199	0.3695	0.84	0.5026
23321	0.186 vs 0.359	0.0826	1.19	0.4894
23333	-0.351 vs -0.067	0.0061*	0.87	0.5477
32211	0.388 vs 0.593	0.0530	1.41	0.2273
32223	0.135 vs 0.361	0.0422*	1.01	0.9994
32232	-0.144 vs 0.070	0.0500*	0.93	0.7464
32313	-0.034 vs 0.089	0.2697	0.89	0.6041
32331	-0.228 vs 0.060	0.0053*	0.92	0.7101
32333	-0.407 vs -0.147	0.0099*	0.73	0.1810
33212	0.220 vs 0.299	0.5244	0.89	0.6494
33232	-0.256 vs -0.044	0.0808	0.73	0.2117
33321	-0.043 vs 0.176	0.1016	0.74	0.2391
33323	-0.274 vs 0.035	0.0098*	0.78	0.3190
33333	-0.479 vs -0.249	0.0013*	0.57	0.0011**

References

Al-Sharifi A, Krynicki C, Upthegrove R (2015) Self-harm and ethnicity: A systematic review. International Journal of Social Psychiatry DOI 10.1177/0020764015573085

Attema A, Versteegh MM, Oppe M, Brouwer WB, Stolk E (2013) Lead time TTO: leading to better health state valuations? Health Economics 22:376–392

Augestad L, Rand-Hendriksen K, Stavem K, Kristiansen I (2013) Time trade-off and attitudes toward euthanasia: implications of using 'death' as an anchor in health state valuation. Quality of Life Research 22:705–714

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

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

Boguszewski R (2012) Zmiany w zakresie wiary i religijności Polaków po śmierci Jana Pawła II. Komunikat z badań. BS/49/2012. Tech. rep., Centrum Badania Opinii Społecznej

Boguszewski R (2015) Kanon wiary Polaków. Komunikat z badań CBOS. Nr 29/2015. Tech. rep., Centrum Badania Opinii Społecznej

Borrill J, Fox P, Roger D (2011) Religion, ethnicity, coping style, and self-reported self-harm in a diverse non-clinical UK population. Mental Health, Religion & Culture 14:259–269

Brooks R, De Charro F (1996) EuroQol: The current state of play. Health Policy 37:53–72

Danyliv A, O'Neill C (2015) Attitudes towards legalising physician provided euthanasia in Britain: The role of religion over time. Social Science & Medicine 128:52–56

Devlin N, Tsuchiya A, Buckingham K, Tilling C (2011) A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Economics 20:348–361

Dolan P, Gudex C, Kind P, Williams A (1996) The time trade-off method: results from a general population study. Health Economics 5:141–154

Ellis L, Wahab E, Ratnasingan M (2013) Religiously and fear of death: a three country comparison. Mental Health, Religion and Culture 16:179–199

Eurobarometer (2005) Social values, Science and Technology. Tech. rep., European Commission, URL `http://ec.europa.eu/public_opinion/archives/ebs/ebs_225_report_en.pdf`

Gallup (2004) Religion in Europe: Trust Not Filling the Pews. Tech. rep., Gallup, URL `http://www.gallup.com/poll/13117/religion-europe-trust-filling-pews.aspx`

Gielen J, Van den Branden S, Broeckaert B (2009) The operationalisation of religion and world view in surveys of nurses' attitudes toward euthanasia and assited suicide. Medicine, health care, and philosophy 12:423–431

Golicki D, Jakubczyk M, Niewada M, Wrona W, Busschbach J (2010) Valuation of EQ-5D Health States in Poland: First TTO-Based Social Value Set in Central and Eastern Europe. Value in Health 13:289–297

Golicki D, Jakubczyk M, Niewada M, Wrona W, Busschbach J (2013) Is extending of a TTO experiment to 23 states per respondent justifiable? An empirical answer from Polish EQ-5D valuation study. Journal of Health Policy & Outcomes Research 1:110–117

Green C, Brazier J, Deverill M (2000) Valuing health-related quality of life. A review of health state valuation techniques. Pharmacoeconomics 17:151–165

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 20:1727–1736

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

Jelsma J, Ferguson G (2004) The determinants of self-reported health-related quality of life in a culturally and socially diverse South African community. Bulletin of the World Health Organization 82:206–212

Matza L, Boye K, Feeny D, Johnston J, Bowman L, Jordan J (2014) Impact of caregiver and parenting status on time trade-off and standard gamble utility scores for health state descriptions. Health and Quality of Life Outcomes 12:48, DOI 10.1186/1477-7525-12-48

Nagpal N, Heid A, Zarit S, Whitlatch C (2015) Religiosity and quality of life: a dyadic perspective of individuals with dementia and their caregivers. Aging & mental health 19

von Neumann J, Morgenstern O (1944) Theory of Games and Economic Behavior. Princeton University Press

Nord E (1992) Methods for quality adjustment of life years. Social Science & Medicine 34:559–569

Nord E (1995) The Person-trade-off Approach to Valuing Health Care Programs. Medical Decision Making 15:201–208

Nord E, Daniels N, Kamlet M (2009) QALYs: Some Challenges. Value in Health 12 Supp. 1:S10–S15

Robinson A, Dolan P, Williams A (1997) Valuing health status using VAS and TTO: what lies behind the numbers? Social Science & Medicine 45:1289–1297

Saffari M, Pakpour A, Naderi M, Koenig H, Baldacchino D, Piper C (2013) Spiritual coping, religiosity and quality of life: A study on Muslim patients undergoing haemodialysis. Nephrology 18:269–275

Seale C (2009) Legalisation of euthanasia or physician-assisted suicide: survey of doctors' attitudes. Palliative Medicine 23:205–212