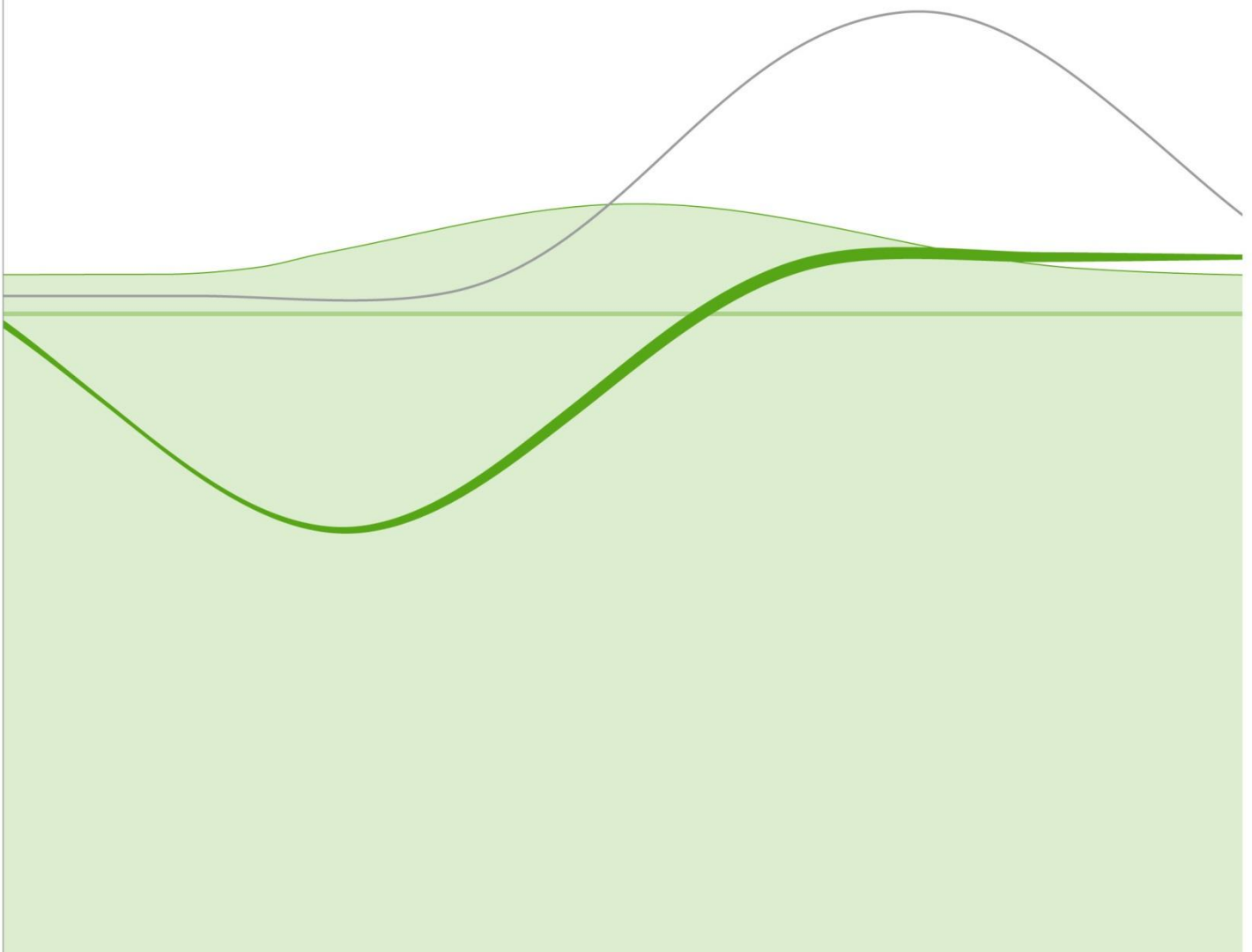# Age and Utilities: Issues for HTA

April 2017

Patricia Cubi-Molla, Koonal Shah, Jamie Garside, Mike Herdman and Nancy Devlin

Age and Utilities: Issues for HTA

Patricia Cubi-Molla[a], Koonal Shah[a], Jamie Garside[b,], Mike Herdman[a] and Nancy Devlin[a]

[a] Office of Health Economics, London
[b] Janssen HEMAR EMEA, High Wycombe, UK

April 2017

For further information please contact:

Patricia Cubi-Molla

PCubi-Molla@ohe.org

The Office of Health Economics

(a company limited by guarantee of registered number 09848965)

Southside, 105 Victoria Street
London SW1E 6QT
United Kingdom
Tel: +44 207 747 8858

About OHE

The Office of Health Economics (OHE) has over 50 years' experience of conducting high quality research on the economics of innovation and the life sciences industry, the organisation and financing of health care, and the role for outcomes research and health technology assessment. OHE is a registered charity in England and Wales (registration number 1170829).

About OHE Research Papers

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication.

Any views expressed are those of the authors and do not necessarily reflect the views or approval of OHE, its Editorial Committee or Research and Policy Committee, or its sponsors.

Once a version of the Research Paper's content is published in a peer reviewed journal, that supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Table of Contents

ABSTRACT

Several studies have found differences in health state values by age. We investigate whether and how age affects respondents' Time Trade-Off (TTO) and Visual Analogue Scale valuations of hypothetical EQ-5D health states using data from the 1993 MVH UK valuation study. Our paper extends upon previous research by the level of detail used in the analysis and by the robustness of the methodology, which minimises the probability of erroneously identifying non-existing differences between age groups (type 1 error). For each profile, the mean TTO or VAS value is pairwise compared across the different age groups. A Bonferroni correction is applied to the multiple testing of significant differences between means. Smile plots are used to illustrate the results. A key finding is the existence of an inverse U-shaped age-utility pattern, with respondents in their forties tending to provide the highest TTO values, and the oldest respondents valuing health profiles systematically (and significantly) lower than younger age groups. This trend is particularly visible for profiles describing problems in the mobility or the self-care dimensions. The paper builds to a discussion on the possibility of using age-specific value sets in health technology assessments, since a technology may not be cost-effective on average but cost-effective for a sub-group whose preferences over health are more closely aligned to the benefits offered by the technology.

1. INTRODUCTION

In England, the National Institute for Health and Care Excellence (NICE) has a clear preference for measuring the health-related quality of life (HRQoL) of patients using the EQ-5D, a generic measure of patient reported outcome, in its health technology assessments (HTAs) (NICE, 2013). A patient's self-reported EQ-5D health state, or profile, can be converted into a utility by assigning the corresponding societal weight from country-specific 'value sets'. In the UK, the value set most commonly used to date for this purpose is referred to as the MVH (Measurement and Valuation of Health) value set (Dolan, 1997). The MVH value set comprises values ranging from -0.59 to 1, where 1 represents full health, 0 represents a health state as bad as "dead", and values less than 0 represent health states considered to be "worse than dead".

These values are used to generate the quality-adjusted life year (QALY) estimates used in cost effectiveness modelling to inform HTA decision making. They are also used to obtain general population reference norms. However, many studies have found differences in the perception, measurement and valuation of health across groups of people of different ages.[1] For instance, important differences in valuation were observed in the time trade-off (TTO) exercise included in the MVH study, where general public survey respondents valued various hypothetical EQ-5D health states (Dolan et al., 1996). Respondents aged 60 years and older assigned lower values than their younger counterparts to approximately half of the health states (specifically, the most severe health states). In addition, the older respondents were found to be more likely to value severe health states as worse than dead. Similar observations have been reported elsewhere – for example, see Wittenberg et al. (2006).

One potential explanation for these differences could be the existence of framing effects linked to the TTO protocol used in the MVH study, especially those effects that are an artefact of the design of the tasks, and may not be observed systematically across age groups. Examples of study design choices that may result in such effects include: using a fixed time frame set equal to 10 years (Van Nooten and Brouwer, 2004; Heintz et al., 2013); offering choices which may be seen as unrealistic by the oldest respondents (Robinson et al., 1997, Witney et al., 2006); and failing to account for the possibility that older people may place relatively less weight on years in the future, as suggested by Dolan et al. (1996). These factors are to some extent controllable in the sense that changes in the structure of the valuation task (e.g. the introduction of a "lead time" in the TTO protocol, as proposed by Robinson and Spencer (2006)) can change the direction and magnitude of these age-sensitive framing effects.

However, even if such effects have been controlled for, individuals of different ages might perceive and interpret the same underlying state of health in different ways. An individual's views about the goodness or badness of a given health state is likely to reflect their aspirations, expectations, fears and priorities, all of which may vary systematically with age. For instance, in the regression model used to construct the MVH tariffs, Dolan (1997) observed age-related differences in the estimated coefficients capturing decrements in utility. For a decrement move from level 1 (no problems) to

---

[1] Evidence suggests that there are also differences between different socio-economic and demographic groups (e.g. Burström et al., 2001). In this paper we will focus on sub-groups defined by age, since it is in turn closely related to health status and therefore has a clear impact on resource allocation.

level 2 (some/moderate problems) respondents aged under 60 years placed greater weight on pain/discomfort, whereas respondents aged 60 years and older placed greater weight on self-care; For a move from level 2 to level 3, pain/discomfort was the most heavily weighted in both age groups). There is additional research supporting these differences: a vignette study conducted by Hofman et al. (2015) shows that older respondents gave significantly more weight to *functional limitations* and *social functioning* and less weight to *morbidities* and *pain experience*, in comparison to younger respondents. However, other authors have found little evidence of systematic differences in societal preferences across age groups (Essink-Bot et al., 2007; Franks et al., 2007).

The potential existence of preference heterogeneity across age groups is not inconsequential. Older individuals are more likely than average to experience multiple co-morbidities, thereby accounting for a large share of health care spending. It is therefore counterintuitive if the general tariffs which are used to inform decisions about the treatment of older people do not reflect their true preferences. This important issue was flagged by Dolan (2000), who proposed the estimation and use of separate valuation tariffs for all EQ-5D health states based on the age group of the respondents.[2] The use of age-specific value sets would open up the possibility that a technology may not be cost-effective on average but cost-effective for a sub-group whose preferences over health are more closely aligned to the benefits offered by the technology (or *vice versa*: cost-effective on average but cost-ineffective for the sub-group).

The aim of this research is to (a) extend the analysis of the relationship between age and utility reported by Dolan et al. (1997) and (b) contribute to the existing debate about the rationale and implications for using age-specific utilities in HTA. We investigate the existence of preference heterogeneity by age group in the MVH value set, that is, whether and how age affects respondents' HRQoL valuations of EQ-5D health states. Our paper extends the previous literature in two main ways. First, we improve the level of detail of the analysis performed by Dolan et al., by introducing six age groups, exploring the utility differences by health profile, and extending the analysis of preference heterogeneity to the VAS reported values. Second, to the best of our knowledge, our paper is the first to apply a robust methodology to identify preference heterogeneity by age group in the MVH value set.

2. DATA AND METHODOLOGY

We use the data collected in 1993 for the MVH study to generate a UK value set for the EQ-5D (MVH Group, 1995). The MVH study elicited preferences from general public respondents regarding a total of 42 hypothetical health states, each described using the EQ-5D classification system and involving at least some health problems (state 11111, which describes no problems on any dimension, was excluded). A selection of 12 EQ-5D profiles was presented to each respondent, with each state valued by 853 respondents on average (minimum: 720). In the first stage of the preference elicitation task, each respondent ranked the 12 EQ-5D profiles, together with 11111, "unconscious" and "dead", having been told to regard these states (other than "dead") as lasting for 10 years followed by death. Next, respondents were asked to locate the health profiles on a

---

[2] It is worth noting that the MVH team reported a large number of TTO- and VAS-based tariffs in their final study report, but only one of those tariffs – the A1 tariff, based on data from the entire sample – was published in the Dolan (1997) paper and is used in practice to generate utility values.

visual analogue scale (VAS), following a bisection process (rating first the best/worst profile, then the one that "comes closest to being half-way on the scale" between the previous two best/worst anchors, and so on). Finally, respondents valued the same 12 health profiles following the TTO protocol, using different methods depending on how the profiles were ranked in the preceding task (i.e. as "better" or "worse" than dead).

The values assigned to hypothetical health states using both methods (VAS and TTO) will be used in this paper. For the purpose of generating data that are comparable across individuals, we follow the MVH approach (1995) and do not use the raw responses (ranging from -39 to 1 for the TTO and from 0 to 100 for the VAS), but rather, values in a re-scaled form with two anchors fixed across individuals. For each method, the reference points are the value given to "dead" (fixed to 0 in TTO, and re-scaled to 0 for VAS) and the value attached to EQ-5D profile 11111 (fixed to 1 in TTO, and re-scaled to 1 for VAS). The TTO values analysed in this paper are those rescaled between -1 and 1, reflecting the approach used by Dolan (1997) to derive the social tariffs. Individual VAS scores for EQ-5D profile k (k from 1 to 12) are adjusted by their relative distance from "dead", using the formula $(VAS_{ik} - d_i)/(11111_i - d_i)$, where VAS_{ik}, d_i and 11111_i stand for the score individual i gives to EQ-5D profile k, "dead" and 11111, respectively. Interval properties are assumed for both measures.

To investigate whether age affects respondents' valuations of EQ-5D health states, we plot the average VAS and TTO values assigned to health states by age group. Age groups are defined based on comparability (taking short age intervals) and sample size requirements (ensuring that each group comprises a sufficient number of observations).[3]

Ordinary least-squares (OLS) regression models are run to estimate the best fitting curves for both measures as a fractional polynomial of age. The analysis is performed for each of the 42 EQ-5D profiles included in the study. However, for simplicity, only the results regarding a sub-sample of profiles is reported in detail. In particular, we select as "mild" profiles those which have no problems in all dimensions except one that has some problems (21111, 12111, 11211, 11121 and 11112); profiles that have severe problems in four or five dimensions are labelled as "severe" (33323 and 33333); and a mixture of profiles with no, some and severe problems in different dimensions is labelled as "moderate" (22222, 21133, 21232 and 12223). For each of these 11 profiles, we plot the average health state values together with the best-fitting curve, by age group.

In order to assess how strongly preferences change across age groups, we test simultaneously for the existence of significant differences between means (null hypothesis: equal means). For each profile, the mean TTO or VAS value is pairwise compared across the different age groups. A Bonferroni correction is applied to the multiple testing, with the aim of minimising the probability of finding significant results by chance. The Bonferroni correction is a rather conservative procedure, potentially leading to a higher rate of "false negatives" (not detecting actual differences between groups) than "false positives" (detecting non-existing differences between age groups), adding robustness to our results (Dmitrienko et al., 2009).

---

[3] It can be criticised that our definition of age intervals is not based on any underlying theory which suggests that the groups are homogenous in their relevant characteristics. We considered alternative age grouping (e.g. "post-war generation" or "baby boomers"), but found no convincing evidence supporting this. As a test of robustness of the results, the analysis was also performed for alternative age cut-off points. No significant differences were found.

"Smile plots" are used to illustrate the results (Newson and ALSPAC Study Team, 2003). These plots represent data points corresponding to every Bonferroni-corrected pairwise test, with the statistical significance (p-value of the test) on the y-axis and the average difference in TTO or VAS values between the corresponding age categories on the x-axis. These plots allow us to readily assess the likelihood of finding significant differences as well as indicating how large those differences might be.

3. HEALTH PREFERENCES BY AGE

From the initial 3,395 respondents, missing, incomplete or unusable data were dropped (MVH Group, 1995, pp. 32-33), giving a final sample size of 2,997 individuals who stated their preferences over different sets of 12 health profiles (a total of 35,964 observations). The sample used in this study is identical to the one originally used to calculate TTO values (Dolan, 1997). A total of 24 observations were dropped from the sample due to missing data for the age of two individuals.[4] Our final sample size is therefore 35,940. The demographic and socio-economic characteristics of the sample are representative of the UK population, and can be found elsewhere (Dolan, 1997).

Table 1 shows some descriptive statistics of the values for the selected health profiles. The minimum number of observations for a profile is 752 (health profile 21133), and the maximum is 2,995 (every respondent had to value profile 33333). For the VAS values we can observe a maximum value of 3.75 for 11112. Scores above 1 correspond to EQ-5D profiles that had been valued as "better than 11111" in the VAS exercise. We can also observe a minimum value as low as -31.33 for 33333. Extremely low values correspond to individuals who assigned close VAS values to "dead" and "full health", and gave severe states a VAS much lower than that given to "dead" (e.g., 95 to "dead", 98 to "full health" and 1 to 33333). Preliminary analysis showed that these extremely low values were affecting the VAS mean in some age groups with a small number of observations, even to the point of showing inconsistent results (one age group valued 33323 as worse than 33333, on average). Two alternative specifications were considered. The first was to use trimmed means, dropping 0.25% of the observations at each side of the distribution. The second was to base all of the analysis corresponding to the VAS on median values rather than on mean values. Both alternatives seemed to correct the results related to the worst health profiles in a similar way. In this paper we present the results corresponding to the trimmed mean values[5]. Any of the alternative specifications are available upon request.

We define 10-year age intervals 18-27 years (label: 22), 28-37 years (label: 32) … 58-67 years (label: 62) and 68 years and above (label: 79). The average number of individuals in each age group (across the different profiles) is 150 (minimum: 90; maximum: 672). Further details about the sample composition are provided in the Appendix (Table A1).

---

[4] MVH Group (1995), Dolan (1997) or Dolan (2000) do not specify how these 24 missing observations are treated.

[5] We drop 90 observations corresponding to 19 individuals reporting VAS adjusted scores below -4.67 (0.25th centile), and 89 observations corresponding to 23 individuals reporting VAS adjusted scores above 1.09 (99.75th centile).

Table 1. Descriptive statistics of TTO and VAS values for selected EQ-5D profiles

	Profile	N	TTO values					VAS values				
			Min	Mean	SD	Median	Max	Min	Mean	SD	Median	Max
Mild	21111	1175	-0.975	0.878	0.226	0.950	1	-2.333	0.793	0.244	0.850	2.216
	12111	1194	-0.975	0.834	0.287	0.925	1	-1.000	0.793	0.179	0.842	1.250
	11211	1209	-0.975	0.869	0.223	0.950	1	-3.000	0.806	0.213	0.850	2.286
	11121	1205	-0.975	0.850	0.242	0.925	1	-2.200	0.807	0.211	0.851	1.250
	11112	1207	-0.975	0.829	0.286	0.925	1	-3.000	0.811	0.235	0.867	3.750
Moderate	22222	770	-0.975	0.500	0.478	0.625	1	-5.857	0.447	0.382	0.500	1.516
	21133	752	-0.975	-0.063	0.594	-0.025	1	-17.000	0.186	0.798	0.250	2.750
	21232	764	-0.975	0.064	0.602	0.138	1	-2.667	0.308	0.329	0.330	1.029
	12223	754	-0.975	0.217	0.559	0.375	1	-7.000	0.319	0.537	0.368	1.408
Severe	33323	761	-0.975	-0.386	0.492	-0.475	1	-26.667	-0.031	1.073	0.070	1.229
	33333	2995	-0.975	-0.543	0.411	-0.625	1	-31.333	-0.135	0.916	0.000	2.143

We investigate the existence of preference heterogeneity (across age groups) based on the respondents' TTO and VAS valuations of health states. For each of the 42 EQ-5D profiles, the corresponding *values* have been adjusted to *age* through quadratic or cubic fitting curves, following OLS regression models.[6] From the results we observe that two thirds of the profiles (28 out of 42) fit better using a quadratic model (TTO values explained as a quadratic function of age, controlling for sex). In these models, the results show a positive coefficient for age and a negative coefficient for age squared, both of which are significant. For every profile, the maximum of the (fitted) TTO values is provided by those aged between 35 and 48 years (this interval narrows down to 42-48 for about 80% of the health states). Only eight of the 42 states (19%) show a better fit in a cubic model. However, the goodness of fit of the quadratic or cubic models for TTO values is quite poor: the best model shows an R squared coefficient equal to 0.05, suggesting that these models can only explain up to 5% of the total variance. Nevertheless, an inverse U-shape can clearly be observed across the different profiles, and is particularly pronounced for some moderate health states. For illustration purposes Figure 1 shows the graph of the best fitted quadratic or cubic equation of the average TTO values given to some mild, moderate and severe health profiles, over the different *age groups.* Note that the dots represented correspond to the actual averages.

Similarly to Figure 1, Figure 2 illustrates the average VAS values in different age groups for the same selection of profiles. From the OLS regressions we observe that only three of the 42 profiles (7%; 11133, 22222 and 23232) fit into the cubic model (showing significant coefficients), with a maximum R squared equal to 0.017, and no profile fits

---

[6] Further details are available from the lead author upon request.

into the quadratic model. In addition, these three profiles do not follow the same pattern. Thus, VAS values seem to be less affected by age than TTO values.

Figure 1. Average *TTO* values by age group, for a selection of (hypothetical) EQ-5D profiles
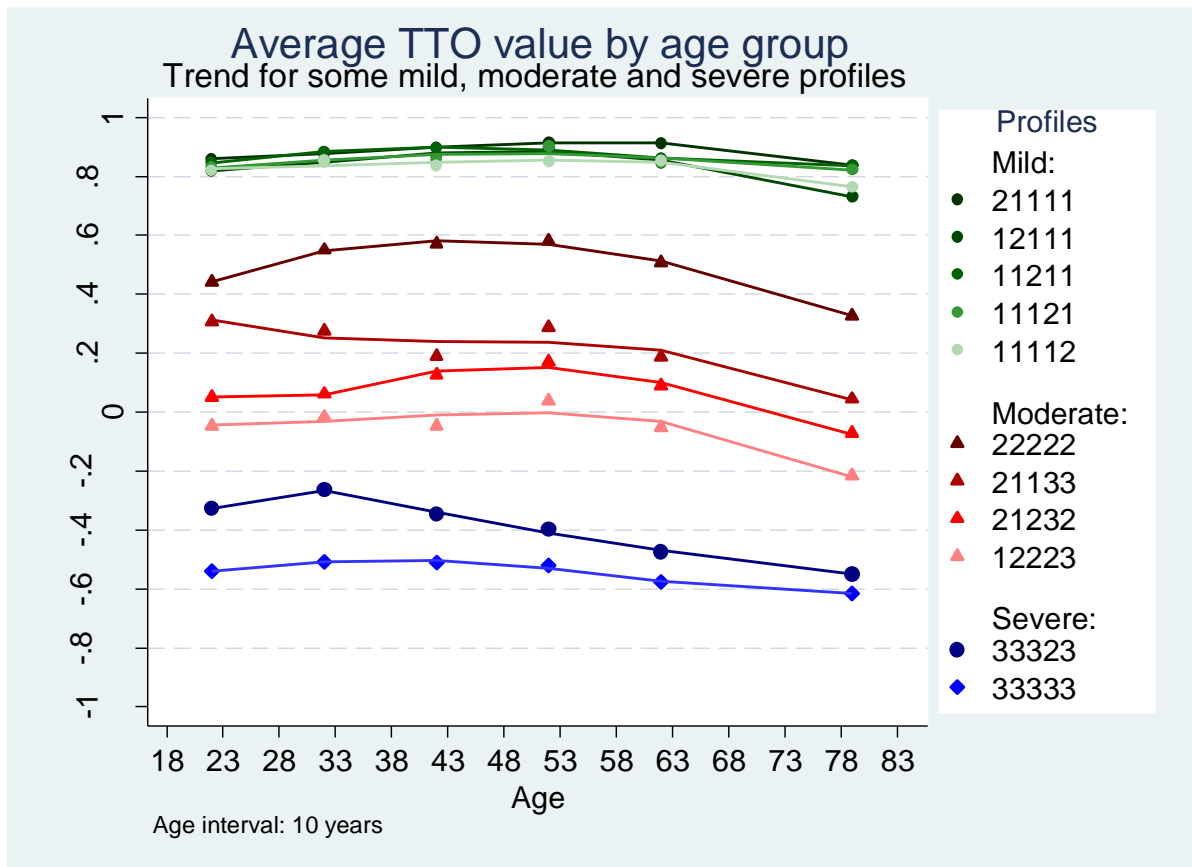
We test for differences in health valuation amongst different age groups. A total of 15 significance tests (pairwise comparison over six age groups) are conducted for each health profile. A Bonferroni correction for the p-values reduces the probability of making at least one false discovery (finding that differences between age groups are significant, when in reality they are not) from 0.54 to the standard 0.05 or 0.01 levels of confidence[7]. Bonferroni test results suggest that the oldest age group (68-90) is the one that most frequently shows differences in TTO values, compared to the average values in other age groups. Of the profiles, 32 out of 42 (76%) show significant differences (at a corrected p-value equal to 0.05) between the TTO valuation of those aged 68 or above and at least one of the lower age groups. Table 2 shows the 20 (out of 42) profiles which also show significant differences at a corrected 0.01 level of significance. These differences are more likely to be observed in moderate and severe health states – they are significant for 11 out of the 13 health profiles with a LSS greater than 11. Only two profiles (33212 and 32211) show significant differences in means between the age group 58-67 and a lower age group.

---

[7] Prob (at least one significant result) = 1 – Prob (no significant results) = 1-(1-0.05)^{15} = 0.54

Figure 2. Average *VAS* values by age group, for a selection of (hypothetical) EQ-5D profiles
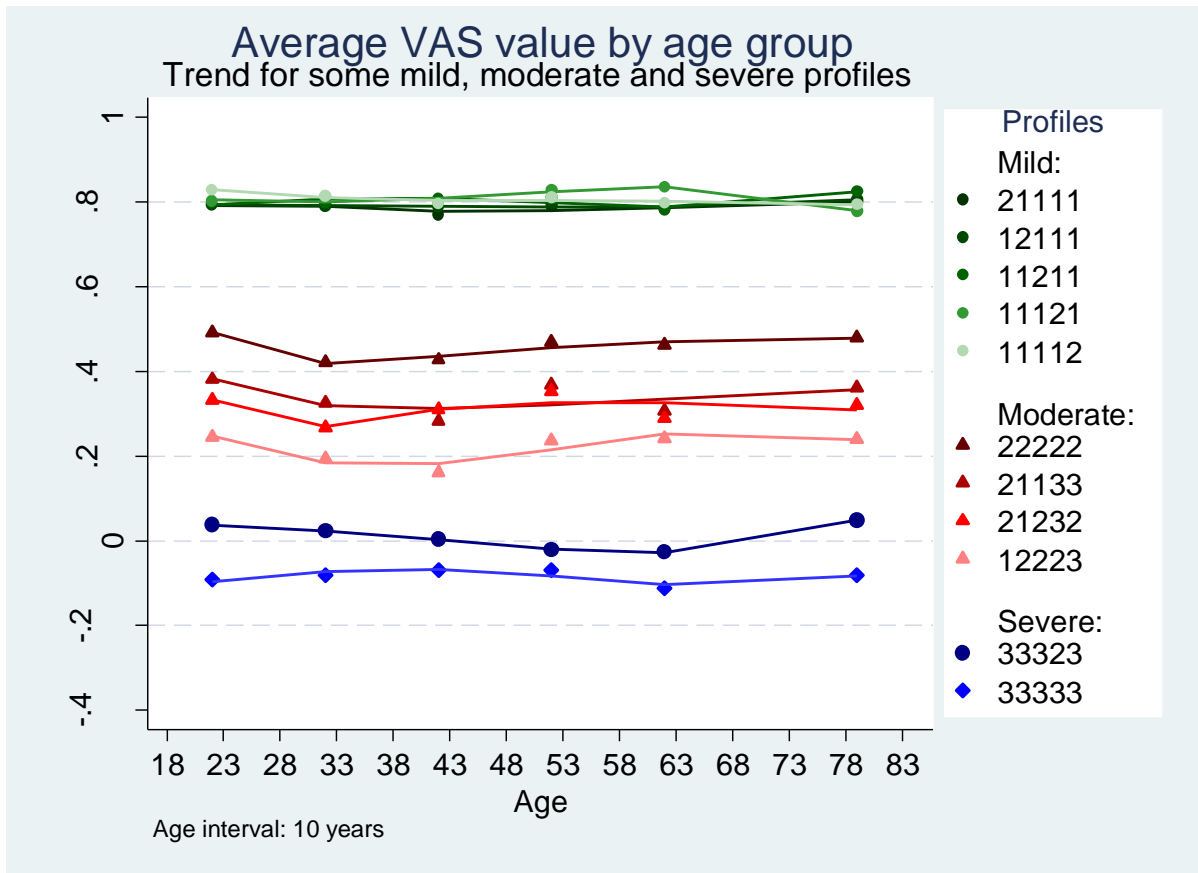

Different probit models were run in order to explore in more detail which dimensions and levels are more associated with the profiles shown in Table 2. We find that significant differences in values amongst age groups seem to be associated with profiles with level 3 in the mobility dimension or level 2 or 3 in the self-care dimension. No significant differences are identified in the remaining dimensions.

The results from the Bonferroni-adjusted tests are illustrated in Figure 3 using a smile plot (Newson and ALSPAC Study Team, 2003). Note that p-values are represented on a log scale. Since most of the significant differences (Table 2) were observed in comparisons involving the age group 68-90, this group is adopted as a reference point. Every small triangle in Figure 3 refers to the results of one of the multiple comparison tests, as indicated in the label. For instance, the upper triangle (labelled "[48-57], 12111") refers to the test comparing the average TTO value attached to profile 12111 obtained from the age groups 48-57 and 68-90 (the null hypothesis is that of equality of the means). The coordinates are, approximately, (0.165, 1.0e-7), which mean that with a probability of almost 1, the average TTO value for profile 12111 obtained from 48-57 year olds is 0.165 points higher than that reported (on average) by 68-90 year olds.

Table 2. EQ-5D profiles showing significant differences (p-value: 0.01) in Bonferroni-adjusted mean value pairwise tests.

Age groups	18-27	28-37	38-47	48-57	58-67	68-90
18-27						
28-37						
38-47						
48-57						
58-67			32211	32211		
		33323				
68-90		12111	12111	12111	12111	
				21111		
		12211	12211	12211	12211	
		13311				
		22122	22122	22122		
		32211	32211	32211		
	12223	12223				
		22222	22222	22222		
	33212	33212	33212	33212		
				13332		
		22323	22323			
			23232			
		23313				
		32223	32223	32223		
			32232	32232		
		32313				
		32331	32331	32331		
	33232					
	33323	33323				
		33333	33333	33333		

Figure 3. Smile plot showing average differences (x-axis) and p-values (y-axis) for every pairwise comparison of mean TTO values amongst age groups, by EQ-5D health profile
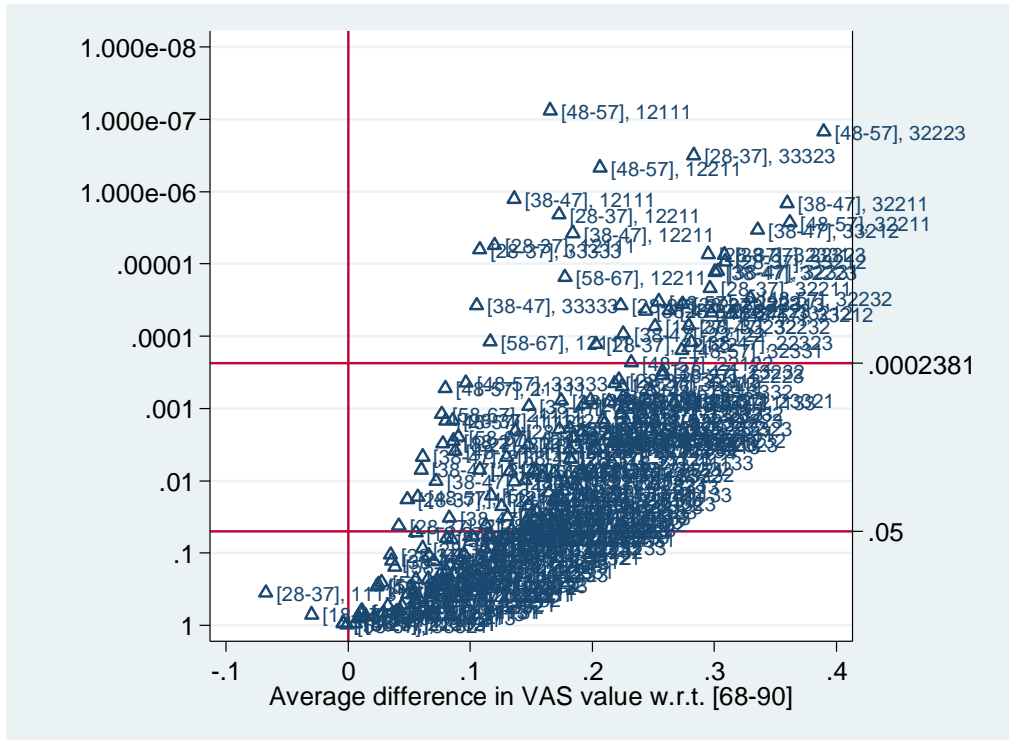
Figure 4. Smile plot showing average differences (x-axis) and p-value (y-axis) for every pairwise comparison of mean VAS values amongst age groups, by EQ health profile
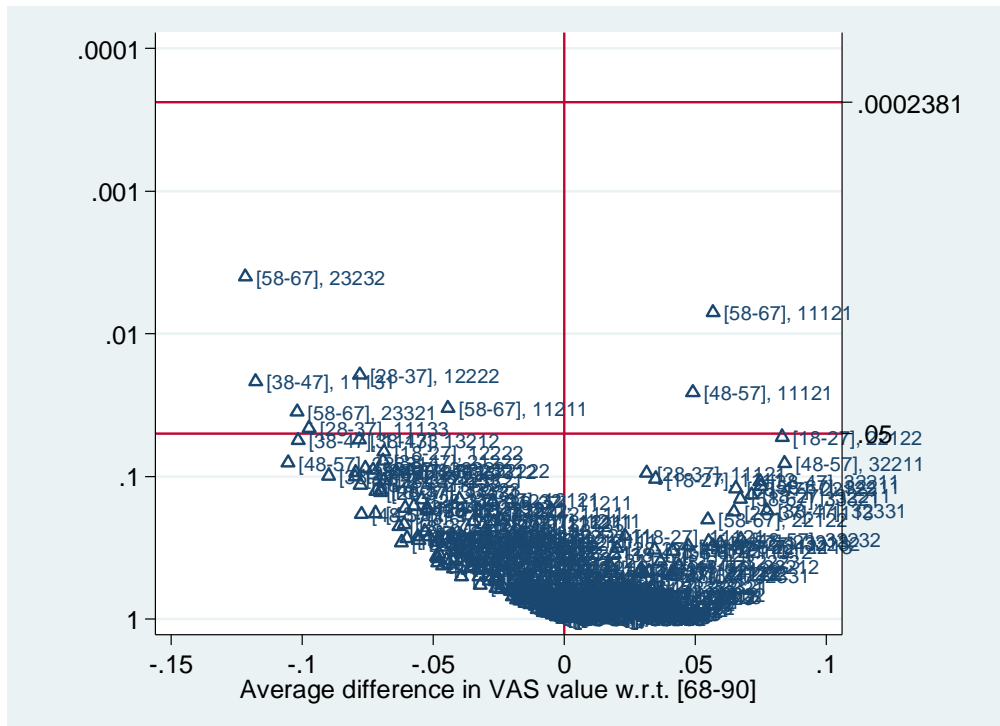
Most of the data points show positive x-values (the x-axis reference line represents 0 expected differences in the values). Thus, we can see that the average TTO values

obtained from the oldest respondents are systematically lower than those obtained from other age groups, with differences ranging between -0.08 (for the age group 28-37) and 0.4. The data points above the lower y-axis reference line represent the pairs of age groups and profiles for which we would have found significant differences, had the uncorrected p-value of 0.05 been used for each test. The upper y-axis reference line ("parapet line", with p-value 0.0002381) depicts the significance level we have to demand in every test, in order to keep the probability of "false positives" (on aggregate) to the order of 0.05. Thus, we could say that with (conservative) 95% confidence, the TTO values stated by the oldest are significantly lower than those stated by the other groups and related to the profiles depicted above the parapet line. The data points between the y-axis reference lines could also show true significant differences amongst age groups, but we cannot separate them from the expected 5% of associations found "by chance".

Bonferroni-corrected multiple comparison tests and smile plots were run for the VAS values. Comparison tests showed no significant differences between average VAS values across age groups. Figure 4 illustrates this fact. We observe that VAS values provided by older people are not systematically higher or lower than those provided by younger age groups. Differences are much smaller (range between -0.2 and 0.2), seem arbitrarily positive or negative, and are not significant under the corrected p-value.

4. DISCUSSION

In this paper we have used the MVH value set data to examine how valuations of hypothetical EQ-5D health states are affected by the ages of the survey respondents who provided the valuations. We find evidence that TTO values are indeed affected by the age of respondents. This study extends upon previous research by the level of detail used in the age-related population groups and in the health profiles. We identify the existence of an inverse U-shaped age-utility pattern, with respondents in their forties tending to provide the highest values for the majority of the health states analysed. To our knowledge this pattern had not been observed before. The TTO values obtained from the oldest respondents are systematically (and significantly, for the majority of profiles) lower than those obtained from younger age groups. The difference between average TTO values obtained from the oldest respondents and those obtained from younger age groups can be as large as 0.4, which can be considered substantial given that a difference of 1 corresponds to the difference between "full health" and "dead". The differences are such that health state 21232 is (on average) considered worse than dead by oldest respondents and better than dead by younger respondent groups.

Our study adds to the literature on determinants of health state values, which has reported mixed evidence on whether and how values are affected by age. Our finding – that age clearly does affect values – can be compared with those of Essink-Bot et al. who found that very little of the variance in health state values was attributable to age (Essink-Bot et al., 2007). The authors instead conclude that the majority of the variance in values observed in their study was due to differences in health states and to individual response patterns. We sought to control for differences in health states by carrying out analyses at the individual profile level (and also for profiles grouped according to their LSS), finding that TTO response patterns *are* related to age. It should be noted a major difference between the studies is that we used the entire MVH data set whereas the data analysed by Essink-Bot et al. was more limited (212 respondents; six health states; three age groups, with very few respondents aged 65 and older). The health states used

in Essink-Bot et al.'s study were also taken from a slightly different classification system, a modified version of EQ-5D that included cognition in addition to the standard five dimensions.

Our study also finds that significant differences in values amongst age groups seem to be associated with profiles with level 3 in the mobility dimension or level 2 or 3 in the self-care dimension. This finding is in line with other studies as in Hofman et al. (2015), where significant differences are found in the weight given to functional limitations and morbidities by different age groups.

A third key finding of our study, however, is that VAS valuations appear to be less affected by age than TTO valuations. This indicates that differences between the values of older and younger respondents may reflect the way that different age groups respond to different preference elicitation techniques. For instance, as highlighted in Robinson et al. (1997), respondents did not seem to be considering the time duration of the health state when choosing a value at the VAS.

TTO involves asking respondents to choose between a shorter life in full health and a longer life in an impaired health state. Hence, a TTO valuation reflects not only how the respondent feels about a given health state, but also how they feel about the trade-off between length of life and quality of life. By contrast, VAS does not involve any explicit trade-off – although it is implicitly stated, since respondents were told before doing the VAS valuations to regard the health states as lasting for 10 years, and also the VAS values were located between 11111 and "dead".

Older respondents and younger respondents may feel differently about the trade-off between longevity and quality of life. For example, younger respondents may place a relatively high value on living longer and would therefore be less willing to give up life expectancy in order to avoid a particular health problem. In addition, older respondents may be less concerned about extending their life and therefore less willing to endure health problems in order to avoid a shortened life. This example suggests that older respondents would tend to give lower values to a given severe health state than younger respondents. This is consistent with the findings of our study. It is also consistent with a systematic literature review (Arnesen and Trommald, 2005) which observes that for TTO self-valuations of the respondent's own current health status, there is a clear trend that the values derived for *the same EQ-5D health self-description* decrease with age.

A related issue is that the TTO variant used in the MVH study asked respondents to imagine that they would live for 10 years, after which they would die. This duration of health state is commonly used in TTO studies and is the duration recommended in the EuroQol Group's protocol for valuation EQ-5D-5L (the new, five-level version of EQ-5D) health states (Oppe et al., 2014). There is evidence that the valuation of a health state is not independent of its duration (Attema and Brouwer, 2010), and older respondents who do not expect to live for a further 10 years might be more willing than average to sacrifice this 'excess' lifetime (Dolan and Roberts, 2002). An alternative interpretation of this result is based on the hypothesis of good innings or less tolerance for themselves, as suggested by Dolan et al. (MVH Group, 1995): some older respondents may feel that they have lived enough, so are less prepared to struggle on in an unsatisfactory health state at the end of their lives.

Reflecting on the MVH study, Dolan (2000) suggests that certain elements of the TTO exercise – specifically, the scenario used to value states considered to be worse than dead – were interpreted differently by respondents aged 60 years and older compared to

those aged between 18 and 59 years. Half of the older respondent group considered this scenario to be implausible, compared to 10% of the younger respondent group. Dolan concludes that at least some of the observed difference between the values of older and younger respondents is likely to be artefactual – "the result of biases introduced by the procedure used to elicit them" (p.20).

4.1. Is There a Case for using Age-Specific Utilities in HTA?

In an ideal world, the preferences of *all* individuals in society could be known and judgements about whether a given technology is cost-effective or not could be made for each person individually. However, this is clearly infeasible, so the average preferences of a sample of the general public are used to infer the preferences of wider society. Sculpher and Gafni (2001) liken this to searching for a figurative 'representative individual', when in fact such an individual cannot exist due to the considerable heterogeneity in people's preferences. Further, the correct way of 'averaging' preference data is unclear (Devlin et al., 2017) and the average of all observed preferences is not the same as the preference of the average person. Sculpher and Gafni (2001) note that the use of average preferences, combined with the decision rule of deeming a technology to be cost-effective if it dominates its comparator or its incremental cost-effectiveness ratio is lower than a given threshold, results in the technology "either being considered 'cost-effective' or 'not cost-effective' *for all individuals* regardless of the variation between individuals which underlies the average preferences" (p.318). Ignoring heterogeneity thus results in a suboptimal use of scarce heath care resources.

Just as clinical sub-group analysis can increase overall health benefits by recognising that the effectiveness of a given treatment can differ across patient sub-groups, (Sculpher and Gafni, 2001) argue that *preference* sub-group analysis can recognise that there may exist a sub-group of a population whose preferences are sufficiently different to the whole-group average so as to produce qualitatively different incremental cost-effectiveness ratios. This opens up the possibility that a technology is cost-*ineffective* on average but cost-*effective* for a sub-group whose preferences over health are more closely aligned to the benefits offered by the technology. The use of preference sub-groups can therefore increase overall health and improve efficiency by making the technology available only for the relevant sub-group, as long as that sub-group can easily be identified (such as one defined by age).

Sculpher and Gafni's proposal has been criticised by Robinson and Parkin (2002), who claim that "using the average values of sub-groups defines these as sub-communities, which […] is only consistent with a separate health service for each of them" (p.650). They argue that sub-group values can reasonably be used to inform decisions being made within a particular clinical context, but not when making "global resource allocation decisions involving community preferences" (p.651).

UK health care decision makers are required to respect anti-discrimination legislation that states that patients must not be denied (or have restricted) access to NHS care because of their race, disability, age, gender, sexual orientation, religion, beliefs or socioeconomic status. This suggests that it would not be acceptable to use age as a basis for defining sub-groups if this results in denying patients access to treatment based solely on their age. On the other hand, NICE's *Social Value Judgements* guide notes that its guidance might be able to refer to age if, amongst other things, "there is good evidence, or good grounds, for believing that because of their age patients will respond

differently to the treatment in question" (NICE, 2008) (p.24). In other words, age-based sub-groups are acceptable if they are *clinically relevant*.

To give an extreme example, suppose there was clear evidence from valuation research that people aged under 60 years placed no weight at all on improvements in self-care whereas people aged 60 years and over placed a great deal of importance on this attribute. This suggests that a treatment whose *only* clinical effect is to improve the patient's ability to wash and dress themselves would generate substantial HRQoL benefits for the older sub-group but no HRQoL benefits for the younger sub-group. These health benefits could be expressed in terms of QALYs – treating older patients would result in a positive QALY gain whereas treating younger patients would not. It is unclear whether such a case would be interpreted by HTA agencies such as NICE as demonstrating relevant differences in effects across age groups, and therefore whether age-specific guidance in relation to this treatment would be justified.

A further complication is that equity issues may arise if an individual patient is denied access to a treatment that they themselves would respond well to (as their own preferences over health are aligned with the benefits offered by the treatment), but they belong to an age-defined sub-group with substantially different preferences on average. Such issues will always arise given that it will never be possible to elicit the preferences of each and every member of society. The sub-groups analyses proposed by Sculpher and Gafni, as Robinson and Parkin (2002) (p.649) put it, are "simply a set of sub-means, with the distribution around them ignored in exactly the same way as a global mean".

A compromise suggested by Dolan (2000) is to use whole population values when assessing treatments that affect multiple sub-groups; but when the decision problem is whether one treatment or another should be recommended for a particular sub-group, there is a case for using the values of the relevant sub-group.

A further issue relates to the way in which QALYs based on utility information are used to inform the appraisal of health technologies. In England, NICE makes recommendations about whether a given technology should be funded by comparing the incremental cost-effectiveness of that technology to a threshold that reflects the opportunity costs. The opportunity costs are the QALYs forgone as a result of displacing existing services, as necessary to fund the new technology. If the estimates of QALYs gained from new technologies are routinely based on age-specific utilities, then this metric will tend to vary on a case-by-case basis (depending on the age groups affected). On the other hand, the estimate of QALYs forgone will always be based on average (non-age-specific) preferences because it is highly unlikely that age-specific utility information about the services forgone for any given decision can ever be known. This asymmetry in the measurement of QALY gains and QALYs forgone may result in an inefficient allocation of health care resources.

4.2. Recommendations for Further Research

Whilst the evidence that TTO values are affected by the age of respondents is strong, further empirical research can provide more insight into the relationship between age and health-related preferences.

The analyses reported in this paper relate to preference data collected in 1993 regarding the three-level EQ-5D instrument. It is likely that the nature of people's preferences will have changed since then (not least because average life expectancy has increased) and it would be worth conducting similar analyses on preference data collected more recently

regarding the new five-level instrument. And since life expectancy changes over time (both at the societal level and for any given individual), the age-utility relationship may also be expected to change. It would be interesting to explore how the preferences of a given cohort of individuals change as they get older.

Finally, this study examined one method that is highly susceptible to age-related artefactual effects (TTO) and another that is less susceptible to such effects (VAS). In recent years, discrete choice experiments (DCEs) have increasingly been used in health state valuation research. DCE studies that require respondents to choose between health states of unspecified duration would generate data that are unlikely to be subject to the same age-related artefactual effects as TTO data, so it would be worthwhile examining the extent to which the age-utility relationships differ between DCE and TTO data.

REFERENCES

Arnesen, T. and Trommald, M., 2005. Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Economics,* 14(1), pp.39–53.

Attema, A. and Brouwer, W., 2010. On the (not so) constant proportional trade-off in TTO. *Quality of Life Research,* 19(4), pp.489-497.

Burström, K., Johannesson, M. and Diderichsen, F., 2001. Swedish Population Health-Related Quality of Life Results Using the EQ-5D. *Quality of Life Research,* 10(7), pp.621-635.

Devlin, N., Shah, K.K. and Buckingham, K., 2017. *What is the normative basis for selecting the measure of 'average' preferences for use in social choices?* Research Paper. London: Office of Health Economics.

Dmitrienko, A., Tamhane, A.C. and Bretz, F. eds., 2009. *Multiple testing problems in pharmaceutical statistics*. In: s.l.:CRC Press.

Dolan, P., 1997. Modeling Valuations for EuroQol Health States. *Medical Care,* 35(11), pp.1095-1108.

Dolan, P., 2000. Effect of age on health state valuations. *Journal of Health Services Research & Policy,* 5(1), pp.17-21.

Dolan, P. and Roberts, J., 2002. To what extent can we explain time trade-off values from other information about respondents? *Social science & medicine*, 54(6), pp.919-929.

Dolan, P., Williams, A., Kind, P. and Gudex, C., 1996. The time trade-off method: results from a general population study. *Health Economics,* 5(2), pp.141-154.

Essink-Bot, M.L., Stuifbergen, M.C., Meerding, W.J., Looman, C.W. and Bonsel, G.J., 2007. Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BMC health services research*, 7(1), p.62.

Franks, P., Lubetkin, E.I. and Melnikow, J., 2007. Do personal and societal preferences differ by socio-demographic group? *Health Economics,* 16(3), pp.319-325.

Heintz, E., Krol, M. and Levin, L., 2013. The impact of patients' subjective life expectancy on time tradeoff valuations. *Medical Decision Making,* 33(2), pp.261-270.

Hofman, C.S., Makai, P., Boter, H., Buurman, B.M., de Craen, A.J., Olde Rikkert, M.G.M., Donders, R. and Melis, R.J., 2015. The influence of age on health valuations: the older olds prefer functional independence while the younger olds prefer less morbidity. *Clinical Interventions in Aging*, 10, pp.1131-1139.

MVH Group, 1995. *The measurement and valuation of health: final report on the modelling of valuation tariffs,* York: Centre for Health Economics.

Newson, R. and the ALSPAC Study Team, 2003. Multiple-test procedures and smile plots. *Stata Journal,* 3(2), pp.109-132.

NICE, 2008. *Social value judgements: Principles for the development of NICE guidance.* Second edition. London: National Institute for Health and Care Excellence

NICE, 2013. *Guide to the methods of technology appraisal,* London: National Institute for Health and Care Excellence.

Oppe, M., Devlin, N.J., van Hout, B., Krabbe, P.F. and de Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, 17(4), pp.445-453.

Robinson, A., Dolan, P. and Williams, A., 1997. Valuing health status using VAS and TTO: what lies behind the numbers? *Social science & medicine*, 45(8), pp.1289-1297.

Robinson, A. and Spencer, A., 2006. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Economics,* 15(4), pp.393–402.

Robinson, A. and Parkin, D., 2002. Recognising diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. A response to Sculpher and Gafni. *Health economics*, 11(7), pp.649-651.

Sculpher, M. and Gafni, A., 2001. Recognizing diversity in public preferences: The use of preference sub-groups in cost-effectiveness analysis. *Health economics*, 10(4), pp.317-324.

Van Nooten, F. and Brouwer, W., 2004. The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Economics,* 13(8), pp.819–823.

Witney, A.G., Treharne, G.J., Tavakoli, M., Lyons, A.C., Vincent, K., Scott, D.L. and Kitas, G.D., 2006. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology*, 45(8), pp.975-981.

Wittenberg, E., Halpern, E., Divi, N., Prosser, L.A., Araki, S.S. and Weeks, J.C., 2006. The effect of age, race and gender on preference scores for hypothetical health states. *Quality of life research*, 15(4), pp.645-653.

APPENDIX

Table A1. frequencies for respondents valuing hypothetical health states, by age group.

	ages	label	Mild					Moderate				Severe	
			21111	12111	11211	11121	11112	12223	21133	21232	22222	33323	33333
Age groups	18-27	22	182	189	175	183	198	119	104	120	110	109	460
	28-37	32	253	278	283	269	261	169	178	179	190	178	672
	38-47	42	203	212	200	191	217	136	131	132	124	133	508
	48-57	52	163	150	153	150	160	99	90	92	112	96	386
	58-67	62	146	158	190	174	170	107	103	103	103	104	420
	68-90	79	219	202	201	235	197	120	143	138	129	138	528