Population-based preference weights for the Adult Social Care Outcomes Toolkit (ASCOT) for Service Users for Austria: findings from a best-worst experiment

Assma Hajji[1], Birgit Trukeschitz[1*], Juliette Malley[2], Laurie Batchelder[3], Eirini Saloniki[3], Ismo Linnosmaa[4,5] and Hui Lu[6]

[1] Research Institute for Economics of Aging, WU Vienna University of Economics and Business, Vienna, Austria

[2] Care Policy and Evaluation Centre, London School of Economics and Political Science, London, UK

[3] Personal Social Services Research Unit, University of Kent, Canterbury, UK

[4] Centre for Health and Social Economics, Finnish institute for health and welfare (THL), Helsinki, Finland

[5] Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland

[6] RAND Europe, Cambridge, UK

*Corresponding author:

Birgit Trukeschitz

Research Institute for Economics of Aging, WU Vienna University of Economics and Business

Welthandelsplatz 1, Building D4

1020 Vienna, Austria

+43-1-313-36-5877

birgit.trukeschitz@wu.ac.at

Declarations of interest: none

Abstract

Background: The Adult Social Care Outcomes Toolkit (ASCOT) measures quality-of-life (QoL) outcomes of long-term care (LTC) service provision. Country-specific preference weights are required to calculate ASCOT scores. ASCOT has been translated into German, but lacks preference weights for German-speaking countries

Objectives: This paper aims to establish Austrian preference weights for the German version of the ASCOT service user measure, using best-worst scaling (BWS).

Methods: Data were collected using an online BWS-experiment from a general population sample (n=1,000) of Austrian adults. We use a scale-adjusted multinomial logit model (S-MNL) accounting for positioning effects to estimate preference weights.

Results: Austrians value the top attribute-levels in the ASCOT domains 'being meaningfully occupied during the day' and 'having control over daily life' most highly, whereas high needs were the least preferred in the domains 'dignity' and 'social participation'. From a methods perspective, we found significant positioning effects only for 'best' choices, with statements at the top of a list being picked more often than those further down in the list. Factors related to survey completion (self-assessed understanding of the tasks and survey completion time) were shown to have the greatest effect on individual choice consistency.

Discussion: The paper provides Austrian preference weights for the German version of ASCOT for service users. The weights also provide insight into how the Austrian general population values different LTC-QoL states in case of care dependency. Future research may investigate how values for different LTC-QoL states differ between socio-economic groups.

Keywords:

ASCOT; best-worst-scaling; preferences; long-term care; quality of life; Austria

1 Background

Quality of life (QoL) is an important outcome indicator for evaluating the effectiveness and cost-effectiveness of health and long-term care (LTC) service provision (Kane, 2001). Economic evaluations of LTC seek to inform policymakers and practitioners about the appropriate allocation of scarce resources in aging populations. Measuring QoL in a LTC context requires concepts reflecting a holistic view of care dependent peoples' lives, such as 'care-related quality of life' (Pieper & Vaarama, 2008) or 'social-care related quality of life' (Netten et al., 2012). In line with this, instruments for economic evaluations have been developed to capture QoL in health and LTC settings, such as the ICECAP index of capability for older people (Coast et al., 2008) or the impact of LTC services on service users' QoL (Makai et al., 2014; Netten et al., 2012). These tools focus on various aspects of people's lives and are therefore better suited for use in LTC service evaluation and decision-making.

The Adult Social Care Outcome Toolkit for Service Users (ASCOT), developed in England, has been translated into several languages (van Leeuwen et al., 2015; Milte et al., 2014; Nakamura-Thomas et al., 2019; Towers et al., 2016) and has recently been translated into German. ASCOT is a LTC-related QoL (LTC-QoL) measure for investigating the impact of LTC services on service users' QoL across eight domains (Netten et al., 2012). These eight domains cover both basic aspects, such as personal cleanliness, food and drink or personal safety, and higher order aspects, such as social participation, occupation or control (Netten et al., 2012). Within each of the eight domains, LTC service users indicate their QoL-state on a 4-level scale, ranging from 'ideal state' to 'high-level needs'. In addition to providing insight into domain-specific outcomes of LTC service provision, domain scores can be combined to generate an overall ASCOT score for the purpose of evaluation.

For economic evaluation, it is important that the outcome measures reflect the value of the outcome state attained. Simply aggregating over all the domains using the same coding system for each domain (e.g. 0, 1, 2, 3) to get a single measure of overall QoL may lead to a biased overall score, for two reasons. First, this assumes that within domains levels are equidistant from each other on the scale. Second, due to the equal coding, this approach implies that comparable states in different domains are valued equally. As has been shown for England (Burge et al., 2010b), people value states in different ASCOT domains of life differently. Thus, economists suggest generating a set of preference weights that reflect the relative importance of the QoL-states represented by the measure for use in economic evaluation.

A validated German version of ASCOT instrument for LTC service users is available for data collection of LTC-related quality-of-life outcomes in German-speaking countries (such as Germany, Austria, Switzerland) and regions (such as South Tyrol in Italy). However, preference weights for the ASCOT Service User instrument are not yet available for any German speaking country (so far, preference weights are only available for England and Japan (Netten et al., 2012, Nakamura-Thomas et al., 2019). As ASCOT has gained interest of policy makers, care organisations and researchers in Austria (Trukeschitz, 2011), the availability of Austrian preference weights would allow for a more accurate utilisation of the translated ASCOT tool in Austria as the data generated using those weights are a better representation of the respondents' LTC-QoL overall than the raw ASCOT scores.

A range of methods is available to develop preference weights (Brazier et al., 2017). Discrete choice experiments (DCE) are quite popular (Carson & Louviere, 2011). Research comparing BWS and DCE in the context of eliciting preferences for the English ASCOT measure, however, found that both methods produce comparable domain-level weights (Potoglou et al., 2011). Additionally, a DCE including all eight ASCOT domains at each of the four levels would either lead to choice tasks that are highly cognitively demanding, thus making the method infeasible for the purpose of this study, or (if using a blocked DCE design with several versions of the questionnaire) require a very large sample size (Witt et al., 2009). Compared to DCEs, best-worst scaling (BWS) experiments make it easier to include more attributes or domains and arguably are less cognitively demanding for respondents (Flynn et al., 2007).

This paper aims to generate population-based preference-weights for Austria for the German version of the ASCOT measure for LTC service users using a BWS experiment. The preference weights reflect how people in Austria value different QoL-states under circumstances in which people have care needs. Policy makers, care managers and researchers can use the results to better understand people's preferences for outcomes from care. In addition, preference weights make it possible to calculate an ASCOT score that reflects the value of the outcome states measured making it suitable for use in the economic evaluations of long-term care services in Austria.

This paper is organised as follows: section 2 lays out the modelling approach, providing details on the choice modelling framework. Section 3 gives insights into the best-worst experiment, the data collection and the sample characteristics. Section 4 presents the results covering the relative frequency of choices and the modelling results. Section 5 discusses the results in light of the aims of the study and the lessons learnt.

2 Modelling preferences using best-worst experiment data

2.1 Theoretical framework and modelling

BWS (Flynn et al., 2007) is a stated-preference method used for eliciting preference weights or utilities based on choices of 'best' or 'worst' (most or least preferred) elements out of choice sets. The best-worst scaling approach is based in random utility theory (RUT) (McFadden, 1974). RUT builds on the assumption that utility is a latent concept, which means that it cannot be observed directly. Utility thus needs to be inferred through a model from variables that are observed. RUT postulates that people choose what they prefer (observed component of the model); any deviation from this (unobserved heterogeneity) is captured by the random component of the utility model (Ben-Akiva et al., 1985; J. Louviere et al., 2002).

The statistical model underlying best-worst scaling uses the relative choice probabilities of a given set of items in order to derive their distance on the latent utility scale. Utilities for all items are therefore estimated on a common scale with a given reference point (often the item with the lowest perceived utility) (Flynn et al., 2007).

Models analysing preferences may account for several sources of heterogeneity (Lancsar et al., 2013; Swait & Bernardino, 2000): taste heterogeneity results from systematic differences in preferences for certain attributes between groups of people (i.e., it is related to observable choices). Another source of heterogeneity, scale heterogeneity, relates to differences in error variance, either between different groups of respondents, tasks or choices (i.e., it is related to the noise in the data and therefore not directly observable) (Davis et al., 2016). Additionally, we consider positioning effects as a third source of heterogeneity potentially affecting choices not relating to the individual or domain itself, but rather to the position the domain was presented at in the task experiment (related to the design of the BWS experiment). They reflect variations in the choice probability of an item given its estimated utility depending on its placement in the list (Campbell & Erdem, 2015; Saloniki et al., 2019)[1].

For the purpose of this paper, we followed the experimental and statistical approach used for eliciting preferences for the ASCOT Service User measure for England (Netten et al., 2012). Additionally, we applied a weighting procedure to address sample non-representativeness where needed, used for the English ASCOT-Carer measure (Batchelder et al., 2019) and previous work by Burge et al. (2010) and Huynh et al (2017). Thus, we ran a covariate-

[1] Since the ordering of domains was randomised across individuals, we did not expect positioning effects to cause preferences to be under- or overestimated. Nevertheless, including them increased the explanatory power of the model significantly and gave more insight into the decision-making process of the respondents.

adjusted (conditional) multinomial logistic model accounting for scale heterogeneity (S-MNL), an extension of the classical multinomial logit model (MNL), which is one of the most frequently used models for analysing BWS data (Mühlbacher et al., 2016). In addition to accounting for scale heterogeneity, the basic model (M1) also considers positioning effects for both best and worst choices, as choice probabilities were affected by the position of the items in the list. Furthermore, we estimated a model (M2) investigating taste differences between subgroups in the sample in order to be able to correct preference weights obtained for sub-groups that were not representative of the Austrian general population. The final preference weights reported for use in economic evaluation were derived from M1, but adjusted for large (>10%) deviations of the sample from the national distribution for those domains or domain levels where significant taste differences were found. Adjustments were applied by post-hoc weighting of affected coefficients from M2 using the correct population proportions. The 10 percentage point cut-off criterion was chosen to avoid over-correction of weights where differences between the sample composition and general population were small and potentially within the margin of error, and followed the approach outlined in Huynh et al (2017) and Burge et al (2010). These coefficients were then rescaled so that total scores lie between 0 and 1 (this follows the approach employed in Burge et al. (2010a)).

2.2 Choice modelling framework

In order to generate the ASCOT preference weights for Austria, we used a scale heterogeneity multinomial logit model (S-MNL) estimating conditional choice probabilities on each of the eight ASCOT domains (henceforth referred to as 'alternatives').

A utility function is set up for each of the alternatives with the following form:

Equation 1: $U_{in} = V_{in} + \varepsilon_{in}$

with V corresponding to the systematic (explainable) part of the utility function and ε representing an error term indicating random variation across both individuals *n* and alternatives *i* (Swait & Louviere, 1993). The systematic part of the utility is determined by a combination of individuals' preferences for alternative levels and additional variables relating to the experiment and respondents.

Assuming that the random term is extreme value distributed, the probability that a given alternative *i* is chosen out of *j* possible alternatives in a choice set *C* is given by:

$$\text{Equation 2: } P(i|C) = \frac{exp(\mu V_i)}{\sum_{j \in C} exp(\mu V_j)}$$

With V_i corresponding to the systematic utility of alternative i. Best and worst choices are modelled simultaneously within the same framework, meaning that the utility coefficient for each alternative when chosen as 'worst' is the negative of the utility of that alternative when chosen as 'best' (Louviere et al., 2015). This is based on the notion that best and worst choices are rooted in the same decision-making processes and can therefore be treated as symmetrical (worst choices are 'reverse-best' choices). μ is a scale parameter allowing for different error variance in specific sub-groups of the sample. It is inversely related to the variance, therefore a higher μ indicates lower error variance and more consistent answers (i.e., choices becoming more deterministic). Separate scale parameters can be estimated for sub-groups of the sample. Differences between the sub-groups can be investigated by assigning μ a value of 1 for one of the subgroups and testing each of the comparison groups for differences from unity (Netten et al., 2012; Swait & Louviere, 1993).

The systematic utility V_i for each of the eight alternatives i is represented as a linear additive function of the products of coefficients for each of the alternative levels and level-specific contrast variables as well as variables related to the positioning of the alternative within the choice task:

$$\text{Equation 3: } V_i = \sum_{i=1}^{8} \sum_{l=1}^{4} \beta_{il} x_{il} + \sum_{p=1}^{8} \beta_{b_p} b_p + \sum_{p=1}^{8} \beta_{w_p} w_p$$

Where l stands for the levels of that alternative, ranging from 1 to 4; and p stands for the position of the alternative in the list, ranging from 1 to 8. The β_{il} coefficients correspond to the individual level utilities and provide the basis for the preference weights. As the x_{il}variables are effects-coded and only one level of an alternative can be chosen at a time, each choice made (best or worst) provides information on one β_{il}. The β_{b_p} $\left(\beta_{w_p}\right)$ coefficients reflect positioning effects for best choices, the b_p (w_p) are dummy variables related to the position of the item that was chosen as best (worst) within the set C.

We tested for taste heterogeneity (differences in preferences between subgroups) by including attribute- and level-group interaction terms as covariates to the basic model. In this paper, we did not aim to explore differences in taste in detail. As we are interested in deriving overall preference-weights for ASCOT levels, we aimed to identify significant taste differences only for groups that were over- or underrepresented in the sample. Where the sample deviates from the national distribution by 10 percentage points or more, insights from the taste heterogeneity analysis were used to correct the affected coefficients to counteract the sample mismatch (in our case, in terms of education and income).

The dependent nature of choices because of repeated measurement (the fact that multiple choices were made by the same person) was taken into account in the final stage of the modelling process only, as this approach led to shorter run times and more flexibility in modelling. Preliminary models were calculated in the software Alogit (2017) and treated all choices as independent observations. Final models were estimated in Biogeme (Bierlaire, 2003) using robust sandwich estimators. The results reported in the following sections of this paper are based on the final model.

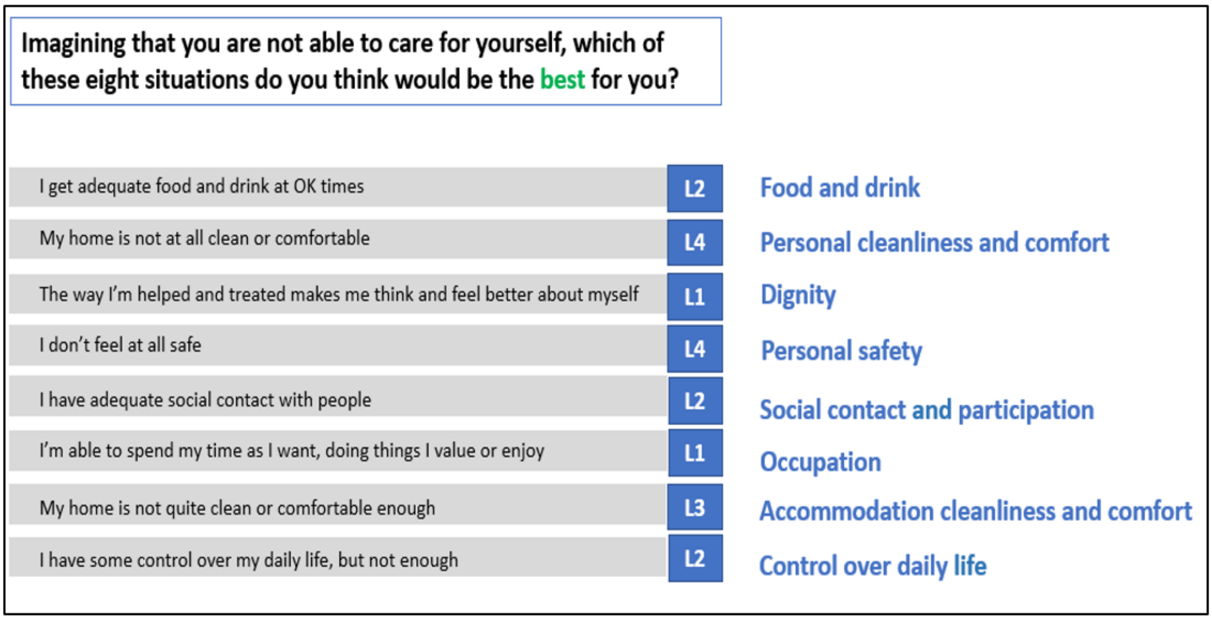
3 Data

3.1 The best worst experiment

Similar to the English approach (Netten et al., 2012), we applied a 'profile case' best-worst experiment (Louviere et al. 2015), which means that the respondents choose attribute levels from a given choice set (or profile). The choice sets were designed using an orthogonal main effects plan (OMEP) (for details see Netten et al. (2012)). The OMEP design relies on the assumption of orthogonality between dimensions that implies a strict linear additivity in the utility functions and allows the estimation of main effects for the eight domains of the ASCOT measure. The full factorial of 4^8 possible scenarios was reduced to 32 scenarios using a fractional-factorial (OMEP) design, which were then blocked into four blocks of eight scenarios to maximise balance and orthogonality. As recommended by Flynn et al. (2010) and Johnson et al. (2013), this base design was checked for profiles with straightforward choices, i.e. profiles where the best or worst choices were very easy to make. In these cases, a foldover version was created which rotated levels within profiles to remove straightforward choices while maintaining orthogonality and balance.

Respondents received eight choice sets, each consisting of a list of statements on eight different aspects of LTC-QoL. They were asked to put themselves in the imaginary scenario of being dependent on other peoples' help due to age-related restrictions, accident or illness. Each statement corresponded to one of four possible levels (response options) of one of the eight ASCOT domains (see Figure 1). The order of domains was randomised across participants, but not within. This was done in order to counteract any positioning effects that may affect choices while keeping the tasks relatively simple for participants (as switching the order of domains within participants might lead to confusion). The levels presented for each domain changed between tasks for each participant according to the experimental plan.

Figure 1: Exemplary illustration of a choice task (choice set)



Note: Levels and descriptions indicating the ASCOT domains were added to Fig.1 for illustration and were not displayed in the survey

Following a step-by-step process, each respondent made four choices per choice set, i.e. a total of 32 choices overall. First, the respondent was asked to choose which item they would consider to be the 'best' (most preferred). The chosen domain was then greyed out and no longer available for the second choice, in which respondents were asked to indicate the least preferred ('worst') statement from the remaining items. This choice was also greyed out and no longer available for the third choice, in which respondents had to select the best situation out of the remaining six statements ('second best'). Finally, from the remaining five statements the respondent chose the situation they least preferred ('second worst'). This resulted in the following choice sequence per choice set: 1st best, 1st worst, 2nd best, and 2nd worst.

3.2 The survey design

The best-worst experiment was part of an online survey consisting of six sections. The first section provided information about the study, gathered the relevant consent to participate and collected data required to monitor the representativeness of the sample. Second, respondents were asked to rate their current quality of life using ASCOT and indicate their overall quality of life. The third section consisted of the best-worst experiment, followed by a section containing questions on the understanding of this experiment. Finally, section five collected information on the respondents' own care experience and section six information on further demographic characteristics.

3.3 Data collection

A representative sample of 1,000 Austrian adults were recruited from a panel (managed by Research Now) to participate in an online survey between August and September 2017. To ensure the representativeness of the sample, quotas were set for sex, age group and living area (all nine Austrian 'Laender') to match official national statistics.

Participants who completed the BWS experiment in an unrealistically short period of time (cut-off criterion = 4.5 minutes) were dropped and sampling following quota requirements continued until the target of 1,000 participants was reached. No further exclusion criterion was applied for the analysis, but we performed sensitivity checks on the results excluding participants who had not fully understood the situations described in the BWS task or those who had not been able to put themselves in the imaginary situations (based on self-assessment questions). Excluding these cases did not significantly alter model results and only slightly improved model fit, hence the full sample was used for the final models.

3.4 Sample

Sociodemographic characteristics of the respondents and corresponding national distributions are given in Table 1.

Table 1: Sample descriptives: gender, age, region, education and income

	Sample		General Population	
	Freq.	Percent	Freq.	Percent
Gender				
Male	473	47.3	3,511,968	48.6
Female	527	52.7	3,708,522	51.4
Age				
18-34 years	297	29.7	1,941,693	26.9
35-54 years	390	39.0	2,554,443	35.4
55 years and over	313	31.3	2,724,354	37.7
Region				
Burgenland	32	3.2	244,753	3.4
Carinthia	66	6.6	468,744	6.5
Lower Austria	196	19.6	1,368,348	19.0
Upper Austria	175	17.5	1,193,948	16.5
Salzburg	55	5.5	449,813	6.2
Styria	155	15.5	1,035,580	14.3
Tyrol	68	6.8	611,991	8.5
Vorarlberg	37	3.7	311288	4.3
Vienna	216	21.6	1,536,025	21.3
Education				
Lower secondary and below	74	7.4	1,644,452	24.6
Upper secondary	510	51	3,675,949	55.0
Short-cycle tertiary and post-secondary	207	20.7	716,501	10.7
Tertiary (BA, MA, PhD or equivalent)	202	20.2	648,530	9.7
Income				
Deciles 1-5	510	60.6		
Deciles 6-10	332	39.4		
Prefer not to say	158			
Total	1000	100.0	6,685,432	100.0

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000), Statistics Austria (http://statistik.at/web_en/statistics/PeopleSociety/population/index.html) and Statistics Austria, EU-SILC 2014

While the distributions for gender, age and region were fairly representative of the national data, educational attainment was not, with those with higher education substantially overrepresented in the sample. Lower income groups (deciles 1-5) were also overrepresented in the sample[2]. Since income and educational attainment deviated substantially (by more than 10 percentage points) from the national population distribution we included these variables in the final model and applied post-hoc corrections to domains or domain-levels where significant taste differences existed for these two variables.

[2] Some information on survey non-respondents (participants who dropped out of the survey) was also collected, namely gender, age and region of the participant. We compared these statistics with the valid sample in order to rule out a systematic bias. We found that non-respondents did not differ from the valid sample in terms of the sociodemographic variables collected.

At the end of the experiment, respondents were asked to rate how well they had understood the choice tasks and if they felt they had been able to put themselves in the hypothetical situations described in the tasks. Table 2 shows the response distributions to these questions. Most respondents had been able to understand the situations all or some of the time, only 1% had not understood the situations at all. In terms of being able to put themselves in the imaginary situations, again almost all of them understood the task; more than half (52%) of the respondents had been able to put themselves in the imaginary situation all of the time and 46% some of the time.

Table 2: Sample descriptives: Understanding of the tasks

In the best-worst exercises, did you understand the situations?		
	Freq.	**Percent**
Yes, all of the time	798	79.8
Yes, but only some of the time	192	19.2
No	10	1.0
Total	1,000	100.0
Did you feel that you could put yourself in the imaginary situations described in the best-worst exercises?		
	Freq.	**Percent**
Yes, all of the time	517	51.7
Yes, but only some of the time	461	46.1
No	22	2.2
Total	1,000	100.0

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000)

4 Results

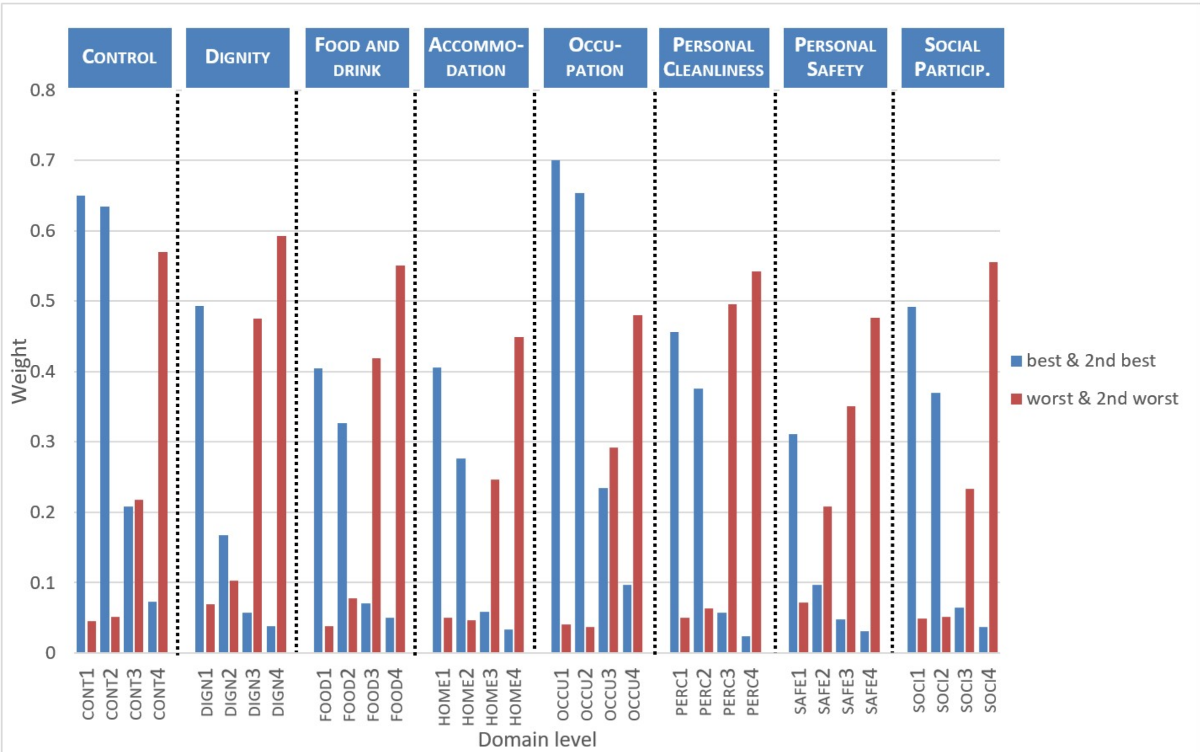
4.1 Descriptive analysis of the choice task data

Figure 2 shows count frequencies of best and worst choices overall. It displays the relative frequency with which an ASCOT item was chosen as best or second best (worst or second worst). Occupation at level 1 (ideal state) was chosen as best (or second best) the most (respondents chose it as best or 2nd best 7 times out of 10 when available). Occupation at level 2 (no needs) and Control at level 1 (ideal state) were chosen as (2nd) best in about 65% of cases. Domains with a high probability of being chosen as preferable at top levels were not necessarily more likely to be chosen as undesirable at bottom levels, indicating that people may wish ‘for’ certain attributes and wish ‘against’ others.

Dignity at level 4 (high needs) was chosen as (2nd) worst the most (in 59% of all possible cases), followed by control (57%), Social participation and Food and Drink (both 55%). Safety at level 2 (no needs) was chosen as (2nd) worst relatively often considering that it relates to an ‘ok’ situation (generally feeling adequately safe, but not as safe as one would like). This indicates that even a minor loss in this domain might be seen as critical and be avoided by respondents.

In rare cases, level 3- or level 4-statements (indicating states with some or high needs) were picked as (2nd) best choices (and, inversely, level 1- or level 2-statements were picked as (2nd) worst in some cases), possibly indicating some error across respondents when performing the choice tasks. This might also be a function of some choice sets, where for example no more level 1 or 2 items were available, or only items in domains that were less important for the participant were available. As some domains (e.g. Safety, Dignity) were more strongly affected than others (e.g. Food and Drink, Occupation), this might also point to some participants reacting very strongly to the overall domain while neglecting the level of the attribute presented in the task.

Figure 2: Descriptive analysis of choice data



Notes: Best and second best (worst and second worst) choices for an item were grouped together, counted over all respondents and choice tasks and divided by the number of times the items appeared. The study used a balanced design, hence this number was equal for all items.

Bars are labelled with the ASCOT domain abbreviation and the level number (1=ideal state, 2=no needs, 3=some needs, 4=high needs).

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000)

4.2 Results of the S-MNL model for Austria

Table 3 shows estimation results for all domain levels from the S-MNL model (M1), accounting for scale heterogeneity and positioning effects. The lowest-rated domain level in the Austrian sample (Dignity 4: 'The way I'm helped and treated completely undermines the way I think and feel about myself') acted as the reference point and was set to zero (Netten et al., 2012). Consequently, all other level-weights were positive.

Table 3: S-MNL estimation results: ASCOT domain level coefficients, scale parameters and positioning effects for Austria

ASCOT: domain levels	Initial coef.	s.e. (rob.)	t-ratio (rob.)[1]	p-value (rob.)	Pairwise significance tests	
					z-ratio	p-value
Accommodation cleanliness and comfort (home)						
1. My home is as clean and comfortable as I want	1.850	0.121	15.289	<0.001	0.316	0.124
2. My home is adequately clean and comfortable	1.660	0.111	14.955	<0.001	7.852	<0.001
3. My home is not quite clean or comfortable enough	0.727	0.057	12.732	<0.001	12.376	<0.001
4. My home is not at all clean or comfortable	0.294	0.037	7.967	<0.001		
Safety (safe)						
1. I feel as safe as I want	1.710	0.113	15.133	<0.001	4.927	<0.001
2. Generally I feel adequately safe, but not as safe as I would like	0.975	0.068	14.317	<0.001	10.936	<0.001
3. I feel less than adequately safe	0.532	0.047	11.247	<0.001	8.179	<0.001
4. I don't feel at all safe	0.289	0.036	8.050	<0.001		
Food and drink (food)						
1. I get all the food and drink I like when I want	1.940	0.127	15.276	<0.001	1.231	0.109
2. I get adequate food and drink at OK times	1.730	0.114	15.175	<0.001	11.003	<0.001
3. I don't always get adequate or timely food and drink	0.391	0.043	9.178	<0.001	5.468	<0.001
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health	0.089	0.035	2.541	0.010		
Personal care (perc)						
1. I feel clean and am able to present myself the way I like	2.010	0.130	15.462	<0.001	1.157	0.155
2. I feel adequately clean and presentable	1.830	0.121	15.124	<0.001	7.474	<0.001
3. I feel less than adequately clean or presentable	0.281	0.038	7.493	<0.001	6.369	<0.001
4. I don't feel at all clean or presentable	0.097	0.034	2.890	<0.001		
Control over daily life (cont)						
1. I have as much control over my daily life as I want	2.450	0.158	15.506	<0.001	0.665	0.376
2. I have adequate control over my daily life	2.380	0.155	15.355	<0.001	8.732	<0.001
3. I have some control over my daily life, but not enough	1.030	0.074	13.844	<0.001	9.073	<0.001
4. I have no control over my daily life	0.015	0.035	0.419	0.680		
Social participation and involvement (soci)						
1. I have as much social contact as I want with people I like	2.100	0.137	15.328	<0.001	1.014	0.077
2. I have adequate social contact with people	1.840	0.120	15.333	<0.001	12.228	<0.001
3. I have some social contact with people, but not enough	0.769	0.058	13.374	<0.001	3.663	<0.001

4. I have little social contact with people and feel socially isolated	0.060	0.033	1.826	0.070		
Dignity (dign)						
1. The way I'm helped and treated makes me think and feel better about myself	2.190	0.141	15.532	<0.001	5.571	<0.001
2. The way I'm helped and treated does not affect the way I think or feel about myself	1.360	0.092	14.751	<0.001	5.343	<0.001
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself	0.283	0.035	8.179	<0.001	4.092	<0.001
4. The way I'm helped and treated completely undermines the way I think and feel about myself	0.000	0.000	-			
Occupation (occu)						
1. I'm able to spend my time as I want, doing things I value or enjoy	2.560	0.164	15.610	<0.001	1.428	0.253
2. I'm able do enough of the things I value or enjoy with my time	2.410	0.155	15.548	<0.001	8.049	<0.001
3. I do some of the things I value or enjoy with my time, but not enough	0.932	0.068	13.706	<0.001	10.723	<0.001
4. I don't do anything I value or enjoy with my time	0.227	0.038	6.037	<0.001		
Scale parameters						
Scale s1: woman	1.190	0.051	3.704	<0.001		
Scale s2: understood only sometimes or not at all	0.653	0.040	-8.763	<0.001		
Scale s3: slower completion (1st quartile[2] to max)	1.720	0.105	6.857	<0.001		
Positioning effects						
post1_B	0.000	0.000	-	<0.001		
post2_B	-0.141	0.021	-6.651	<0.001		
post3_B	-0.191	0.022	-8.843	<0.001		
post4_B	-0.211	0.022	-9.679	<0.001		
post5_B	-0.264	0.022	-11.839	<0.001		
post6_B	-0.296	0.023	-13.040	<0.001		
post7_B	-0.319	0.023	-13.870	<0.001		
post8_B	-0.324	0.023	-13.906	<0.001		
Observations	32,000					
Adjusted rho-squared	0.235					

Notes: [1] *t-ratios for scale parameters were originally calculated with respect to 1 and adjusted accordingly in the table. All included scale factors remain significant after controlling for this.*

[2] *1st quartile threshold: 7.2 minutes*

Note that these coefficients are not corrected for sample non-representativeness, the corrected final weights are reported in chapter 5.4.

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000), own calculations

As the coefficients represented relative values assigned to LTC-QoL-states, we expected them to be highest in top domain-levels and decrease monotonically (level 1 should have a higher coefficient than level 2, and so on, within a given domain). Most coefficients were significantly different from the reference category Dignity 4 ('The way I'm helped and treated completely undermines the way I think and feel about myself '), with the exception of Control 4 ('I have no control over my daily life') (p-value: 0.68) and Social 4 ('I have little social contact with people and feel socially isolated') (p-value: 0.07), which were not significantly different at the 0.05 level. Statistically, this means that these two LTC-QoL attribute-levels were valued as being equally bad as Dignity 4.

The highest-rated attribute-level was Occupation 1 ('I'm able to spend my time as I want, doing things I value or enjoy'), followed by Control 1 ('I have as much control over my daily life as I want') and Occupation 2 ('I'm able do enough of the things I value or enjoy with my time').

The last two columns report z-ratios and p-values for significance tests between a coefficient and the coefficient for the attribute-level below it. In some cases (for the domains Control, Food and drink, Accommodation cleanliness, Occupation, Personal cleanliness and Social participation) the top two levels were not significantly different from each other (p-values: 0.37 (control), 0.11 (food), 0.12 (accommodation), 0.25 (occupation), 0.16 (personal care) and 0.08 (social participation)).

As expected, the coefficients decreased in magnitude within domains, with level 1 (ideal state) having the highest value and level 4 (high needs) having the lowest. For most domains, we see a large difference when moving from level two to level three, indicating that a move from an 'OK' state (level 2) to a state with some unmet needs (level 3) was associated with a considerable perceived loss of utility. Differences between level 1 and level 2 and those between level 3 and level 4 were mostly smaller. Only the Safety domain showed fairly equidistant spaces and significant differences between all levels (all p-values <0.001). Looking at the weights of the top (ideal) levels (level 1), we see that ideal states were valued differently across domains. An ideal state in the area of Occupation or Control was valued very highly, followed by ideal states relating to Dignity and Social interaction. Interestingly, even level two-states (an 'OK' situation) in the domains of Occupation and Control were valued higher than all other level one-states.

Scale heterogeneity analysis: Several tests for scale heterogeneity were performed on the basic model for different sup-groups (based on gender, age, education, task completion time, task understanding, type of choice (best/worst), and difficulty of choice set). The final model included scale parameters for gender, speed and understanding of the tasks, as this proved to be the best specification in terms of model fit. Scale parameters are set to one for a reference group and estimated for the other group. The higher a scale parameter for a certain group, the more consistent the choices are relative to the reference group. Conversely, a lower value indicates higher variance and less deterministic choices.

Scale parameters are shown in table 3. Choices were more consistent in women (compared to men), people who had spent more time completing the tasks (compared to persons who were faster) and people who stated they understood the situations described in the choice tasks (compared to those who didn't understand at all or only some of the time) (all p-values <0.001)[3].

The role of positioning effects: We controlled for potential positioning effects related to the order of the statements presented in the experiment. As positioning effects were expected to affect best and worst choices differently (Campbell & Erdem, 2015), we included coefficients for both best and worst choices at every position in the decision set. Table 3 shows the estimation results, standard errors and t-ratios for the positioning variables for best choices from model 1. Results show that items were less likely to be picked as best choices the further down they were in the list (compared to their chance of being picked as best choices if they were first in the list). All positioning coefficients for best choices were highly significant, but this was not the case for worst choices. Thus, only positioning effects for best choices were included in the model.

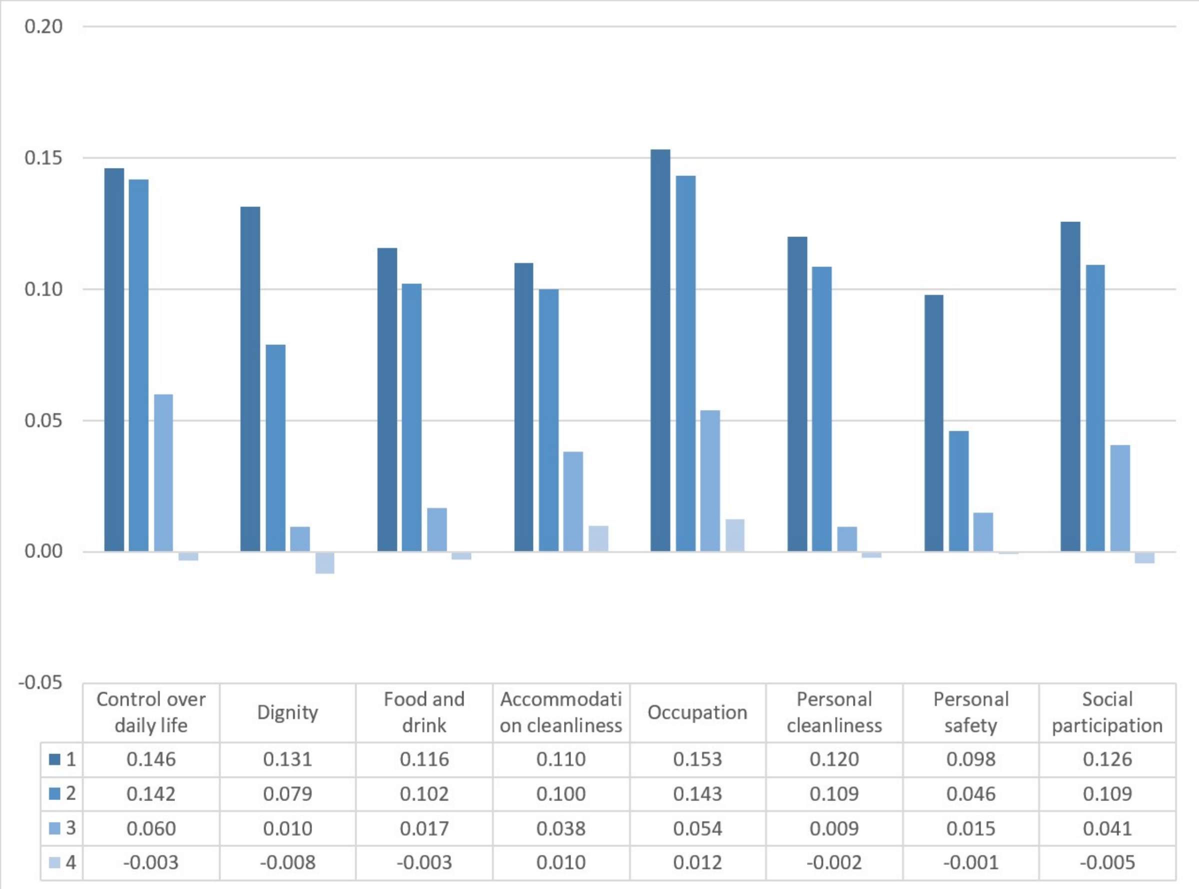
4.3 Final preference weights for the ASCOT Service User measure for Austria

Figure 3 shows the final Austrian preference weights to be used with the German ASCOT Service User measure (adjusted coefficients from M1). The weights were obtained by correcting coefficients that varied by income and education (as described in section 2.1) and rescaling all coefficients in such a way that the range of possible overall LTC-QoL scores (i.e., the sum score over all dimensions) was between zero and one[4]. Results from the model including taste coefficients are shown in Appendix 1. Preference-weighted total scores for individuals are obtained by assessing their LTC-QoL using the ASCOT instrument and summing up level-specific preference weights (as provided in Figure 3) over all LTC-QoL domains (for further details on score calculation see Netten et al. (2012))

Figure 3: Final preference weights of the ASCOT Service User measure for Austria

[3] We also tested whether scale parameters for best and worst choices made in the tasks differed significantly. We found small differences (higher variance for 'worst' choices), which were however not substantial enough to be included in the final S-MNL.

[4] This was done by subtracting 1/8 of the lowest possible sum score from each coefficient and dividing it by the range (highest possible sum score – lowest possible sum score)



	Control over daily life	Dignity	Food and drink	Accommodation cleanliness	Occupation	Personal cleanliness	Personal safety	Social participation
1	0.146	0.131	0.116	0.110	0.153	0.120	0.098	0.126
2	0.142	0.079	0.102	0.100	0.143	0.109	0.046	0.109
3	0.060	0.010	0.017	0.038	0.054	0.009	0.015	0.041
4	-0.003	-0.008	-0.003	0.010	0.012	-0.002	-0.001	-0.005

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000), own calculations

5 Discussion

This paper presents a set of Austrian preference weights to use with the German version of the ASCOT Service User measure. The Austrian population values ideal attribute-levels in higher order LTC-QoL domains (e.g. ‘being meaningfully occupied during the day’ or ‘having control over daily life’) relatively higher than those in basic domains (e.g. ‘feeling clean’, ‘clean and comfortable accommodation’ or ‘feeling safe’). The Austrian sample particularly sought to avoid QoL-states that undermine dignity, and QoL-states that reflect loneliness, no control over daily life and inappropriate food and drink. This is in line with results from the English study (Netten et al., 2012), where ideal states in the domains of Control over daily life and Occupation were favoured most. The top four highest rated states (Control at level 1 and 2 and Occupation at level 1 and 2) were the same in both countries, only the ordering differed. There was some overlap in the lowest-ranked states, with English results showing that high-needs states in the domains of Control over daily life, Safety, Occupation and Food and Drink were seen as particularly undesirable.

The results of the Austrian preference study further show for nearly all ASCOT domains - excluding safety – that the assumption of equidistant spaces between levels cannot be upheld, illustrating the importance of using preference weights with ASCOT. The steep drop in perceived utility when moving from level two to level three indicates that a major change in utility only incurs past a certain threshold, i.e. when some unmet needs arise. In comparison, a move from an ideal to an OK situation does not inflict a similarly high loss of utility. Austrian study participants valued ideal situations in the eight domains of the ASCOT measure differently, placing the highest value on ideal attribute-levels in the domains of occupation and control over daily life and comparatively lower values on ideal attribute-levels in personal safety and accommodation cleanliness. In contrast, situations associated with high needs are seen as especially bad when occurring in the domains of dignity or social participation. Thus, preference-weights account for differences in values across QoL-states and enable the calculation of a single ASCOT score that reflects these differences.

In terms of design effects due to the best-worst experiment, we found a significant positioning effect only for 'best' choices, where respondents were less likely to pick an item as 'best' or 'second best' the further down it was in the list. No pattern was found, however, for 'worst' and 'second worst' choices. This is consistent with findings by Campbell and Erdem (2015) and emphasizes the importance of (i) randomizing the domain order between participants to avoid decision heuristics distorting estimation results, in particular when it comes to 'best' choices and (ii) including positioning effects for best choices in statistical models in order to reduce noise and improve the estimation.

The S-MNL model showed significant differences in consistency of responses across several groups in the Austrian sample. Both 'speeders' and individuals who had not fully understood the tasks were less consistent than those who took their time and felt they understood the exercises, respectively. As scale heterogeneity is related to choice consistency, these results indicate the importance of clear instructions and understandability of the tasks presented.

The present study has some limitations. First, compared to face-to-face interviews, the online survey meant it was not possible to control for certain respondent behaviours during the experiment, such as the attention given to the experiment. The quality of the answers given hinges on the assumption that respondents were able to put themselves in the imaginary situation of being in need of care when completing the exercises. Although we tried to address this with cognitive interviews using the think-aloud method prior to the field phase, by instructing participants before starting the online best-worst scaling exercises and by using prompts to repeatedly remind participants of it throughout the online questionnaire, we cannot be completely sure that online survey participants were considering this imaginary situation when giving their choices. Some assurance is however found in the results of a recent study comparing the results of best-worst scaling experiments using an online survey with face-to-face data collection which showed no notable differences in estimates (Saloniki et al., 2019).

A second issue relating to study design is the use of a general population sample instead of a sample of care-dependent persons themselves. The choice between social preferences (based on general population weights) versus experience-based preferences (based on affected population weights) is not a trivial one and has been subject to debate. While it may be the case that preferences elicited based on a general population sample underestimate the negative effects of certain restrictions, LTC-QoL preferences obtained from the affected population may be biased in the other direction for reasons such as lowered expectations and response shifts. Furthermore, public preferences are generally viewed to be more suitable when dealing with public (health) care expenditures as is generally the case in LTC expenditures (Jonker et al., 2017; Ubel et al., 2003). We also checked whether preferences differed in persons whose QoL in the ASCOT domains was restricted (the ASCOT measure was included in the questionnaire and we performed heterogeneity tests for groups differing in LTC-QoL), but found no meaningful differences. This is in line with results from the English study, which compared preferences obtained from a service user sample with preferences obtained from a general population sample and found no meaningful differences on average (Netten et al., 2012).

Third, as the aim of the current paper was to establish overall Austrian preference weights for the ASCOT instrument for service users, we did not look further into the issue of taste heterogeneity between groups. Future work could explore how perceptions of the value of ASCOT QoL statest are formed or, alternatively, differ between groups of people by assessing taste heterogeneity in more detail. Moreover, as ASCOT preference weights become available for different countries, it would be interesting to understand how preferences for QoL-states might differ across different cultures and regions.

The preference weights for the ASCOT Service User measure presented in this paper can be used for (economic) evaluations of LTC services provided in Austria. Such analyses could give further insight into the benefits and unmet needs of care-dependent service users and may help Austrian managers and policy makers to make evidence-based decisions about the use of resources.

6 Disclaimer

The ASCOT measure is disclosed in full herein but ordinarily should not be used for any purposes without the appropriate permissions of the ASCOT team and the copyright holder – the University of Kent. Please visit www.pssru.ac.uk/ascot or email ascot@kent.ac.uk to enquire about permissions.

7 Acknowledgements

This publication reports on results of the EXCELC (Exploring Comparative Effectiveness and Efficiency of Long-term Care) project, which was funded by the NORFACE Welfare State Futures programme under grant number 462-14-160. In addition, the Austrian contribution to this project was co-funded by the Austrian Science Fund (FWF, project number I 2252-G16) and the Vienna Social Fund (Fonds Soziales Wien). The views expressed are not necessarily those of the funders. The authors would like to thank all participants for taking part in the current study, as well as Research Now, the organisation responsible for data collection. We particularly thank Peter Burge (Rand Europe) for his valuable and constructive suggestions during the planning and development of the best-worst experiment and and Julien Forder (University of Kent) for the acquisition of funding and general oversight of the EXCELC project. The OPEN ACCESS publication of this article was jointly funded by WU Vienna University of Economics and Business and its Research Institute for Economics of Aging.

8 Appendix 1

Table 4-Appendix provides the results of the taste heterogeneity analysis, which have only been used for correcting the population weights.

Table 4-Appendix: MNL estimation results with taste heterogeneity: group-specific ASCOT domain level coefficients

ASCOT: domain levels	**Initial coef.**	**s.e. (robust)**	**t-ratio (robust)[1]**
Accommodation cleanliness and comfort (home)			
1. My home is as clean and comfortable as I want	1.860	0.120	15.510
2. My home is adequately clean and comfortable (Education: lower secondary and below)	1.870	0.163	11.520
2. My home is adequately clean and comfortable (Education: upper secondary/short-cycle tertiary and post-secondary/tertiary)	1.650	0.109	15.120
3. My home is not quite clean or comfortable enough	0.728	0.057	12.870
4. My home is not at all clean or comfortable (Education: lower secondary and below/upper secondary/short-cycle tertiary and post-secondary)	0.282	0.038	7.430
4. My home is not at all clean or comfortable (Education: tertiary)	0.349	0.061	5.690
Safety (safe)			
1. I feel as safe as I want (Income: deciles 1-5)	1.770	0.116	15.260
2. Generally I feel adequately safe, but not as safe as I would like (Income: deciles 1-5)	0.956	0.067	14.270
3. I feel less than adequately safe (Income: deciles 1-5)	0.466	0.046	10.210
4. I don’t feel at all safe (Income: deciles 1-5)	0.217	0.036	6.120
Safety, all levels (Income: deciles 6-10)	-0.199	0.032	-6.130
Food and drink (food)			
1. I get all the food and drink I like when I want	1.950	0.126	15.460
2. I get adequate food and drink at OK times	1.740	0.113	15.380

3. I don’t always get adequate or timely food and drink	0.395	0.043	9.230
4. I don’t always get adequate or timely food and drink, and I think there is a risk to my health	0.086	0.035	2.420
Personal care (perc)			
1. I feel clean and am able to present myself the way I like	2.020	0.129	15.690
2. I feel adequately clean and presentable	1.840	0.119	15.390
3. I feel less than adequately clean or presentable	0.280	0.038	7.440
4. I don’t feel at all clean or presentable	0.095	0.034	2.800
Control over daily life (cont)			
1. I have as much control over my daily life as I want	2.410	0.153	15.770
2. I have adequate control over my daily life	2.340	0.150	15.620
3. I have some control over my daily life, but not enough	1.050	0.073	14.260
4. I have no control over my daily life (Education: lower secondary and below/upper secondary/short-cycle tertiary and post-secondary)	0.055	0.035	1.550
4. I have no control over my daily life (Education: tertiary)	0.224	0.049	4.600
Social participation and involvement (soci)			
1. I have as much social contact as I want with people I like	2.110	0.136	15.530
2. I have adequate social contact with people	1.850	0.119	15.530
3. I have some social contact with people, but not enough	0.772	0.057	13.530
4. I have little social contact with people and feel socially isolated	0.059	0.033	1.800
Dignity (dign)			
1. The way I’m helped and treated makes me think and feel better about myself	2.200	0.139	15.800
2. The way I’m helped and treated does not affect the way I think or feel about myself	1.370	0.091	15.020
3. The way I’m helped and treated sometimes undermines the way I think and feel about myself	0.282	0.035	8.140
4. The way I’m helped and treated completely undermines the way I think and feel about myself	0.000		
Occupation (occu)			
1. I’m able to spend my time as I want, doing things I value or enjoy	2.510	0.158	15.870
2. I’m able do enough of the things I value or enjoy with my time	2.350	0.148	15.850
3. I do some of the things I value or enjoy with my time, but not enough	0.943	0.067	14.110
4. I don’t do anything I value or enjoy with my time (Education: lower secondary and below/upper secondary/short-cycle tertiary and post-secondary)	0.290	0.040	7.250
4. I don’t do anything I value or enjoy with my time (Education: tertiary)	0.346	0.046	7.590
Scale parameters[1]			
Scale s1: woman	1.190	0.050	23.730
Scale s2: understood only sometimes or not at all	0.658	0.039	16.720
Scale s3: slower completion: 1st quartile[2] to max	1.710	0.102	16.700
Positioning effects			
post1_B	0.000		
post2_B	-0.142	0.031	-4.520
post3_B	-0.187	0.033	-5.760
post4_B	-0.211	0.034	-6.200
post5_B	-0.266	0.036	-7.420
post6_B	-0.296	0.037	-7.950
post7_B	-0.324	0.040	-8.170
post8_B	-0.329	0.041	-8.060
Observations	32,000		
Adjusted rho-squared	0.237		

Notes: Significant taste interactions were considered for groups with a <10% difference from general population distributions (education, income)

Coefficients with taste heterogeneity are group-specific: terms in parentheses indicate the group the reported values apply to. All other coefficients (no parentheses) are not group-specific.

[1] t-ratios for scale parameters were originally calculated with respect to 1 and adjusted accordingly in the table. All included scale factors remain significant after controlling for this.

[2] 1st quartile threshold: 7.2 minutes

Source: WU, EXCELC B/W-SU AUT 2017 (n=1,000), own calculations

9 Literature

ALOGIT Software & Analysis Ltd (2017). ALOGIT version 4.2. Retrieved from http://www.alogit.com/

Batchelder, L., Malley, J., Burge, P., Lu, H., Saloniki, E. C., Linnosmaa, I., & Forder, J. (2019). Carer Social Care-Related Quality of Life Outcomes: Estimating English Preference Weights for the Adult Social Care Outcomes Toolkit for Carers. Value in Health, 22(12), 1427-1440.Ben-Akiva, M.E., Lerman, S.R., & Lerman, S.R. (1985). Discrete choice analysis: theory and application to travel demand: MIT press.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models.

Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). Measuring and valuing health benefits for economic evaluation: OXFORD university press.

Burge, P., Netten, A., & Gallo, F. (2010). Estimating the value of social care. Journal of Health Economics, 29, 883-894.

Burge P, Potoglou D, Kim CW, Hess S. (2010b). How do the public value different outcomes of social care? Estimation of preference weights for ASCOT.

Campbell, D., & Erdem, S. (2015). Position bias in best-worst scaling surveys: a case study on trust in institutions. American Journal of Agricultural Economics, 97(2), 526-545.

Carson, R.T., Louviere, J.J. (2011). A common nomenclature for stated preference elicitation approaches. Environmental and Resource Economics, 49(4), 539-559.

Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J., et al. (2008). Valuing the ICECAP capability index for older people. Soc Sci Med, 67.

Davis, K.J., Burton, M., & Kragt, M.E. (2016). Discrete choice models: scale heterogeneity and why it matters. Working Paper 1602: School of Agricultural and Resource Economics, University of Western Australia, Crawley, Australia.

Flynn, T.N., Louviere, J.J., Peters, T.J., & Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. Journal of Health Economics, 26, 171-189.

Flynn, T.N., Louviere, J.J., Peters, T.J., & Coast, J. (2010). Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. Social Science & Medicine, 70, 1957-1965.

Huynh, E., Coast, J., Rose, J., Kinghorn, P., and Flynn, T. (2017). Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. Soc Sci Med.; 189: 114–128

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., ... & Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value in Health, 16(1), 3-13.

Jonker, M.F., Attema, A.E., Donkers, B., Stolk, E.A., & Versteegh, M.M. (2017). Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. Health economics, 26(12), 1534-1547.

Kane, R.A. (2001). Long-Term Care and a Good Quality of Life: Bringing Them Closer Together. The Gerontologist, 41, 293-304.

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., & Burgess, L. (2013). Best worst discrete choice experiments in health: Methods and an application. Social Science & Medicine, 76, 74-82.

Louviere, J., Street, D., Carson, R., Ainslie, A., Deshazo, J., Cameron, T., et al. (2002). Dissecting the random component of utility. Marketing letters, 13(3), 177-193.

Louviere, J.J., Flynn, T.N., & Marley, A.A.J. (2015). Best-worst scaling: Theory, methods and applications: Cambridge University Press.

Makai, P., Brouwer, W.B., Koopmanschap, M.A., Stolk, E.A., & Nieboer, A.P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. Soc Sci Med, 102.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers in Econometrics pp. 105-142). New York: Academic Press.

Milte, C. M., Walker, R., Luszcz, M. A., Lancsar, E., Kaambwa, B., & Ratcliffe, J. (2014). How important is health status in defining quality of life for older people? An exploratory study of the views of older South Australians. Applied health economics and health policy, 12(1), 73-84.

Mühlbacher, A.C., Kaczynski, A., Zweifel, P., & Johnson, F.R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health economics review, 6(1), 2.

Nakamura-Thomas, H., Morikawa, M., Moriyama, Y., Shiroiwa, T., Kyougoku, M., Razik, K., & Malley, J. (2019). Japanese translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT) in Japanese social service users. Health and quality of life outcomes, 17(1), 59.

Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.M., Brazier, J., et al. (2012). Outcomes of social care for adults: developing a preference-weighted measure. Health Technology Assessment, 16, 1-166.

Pieper, R., Vaarama, M. (2008). The concept of care-related quality of life. Care-related quality of life in old age: Concepts, models and empirical findings, 65-101.

Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., et al. (2011). Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Social Science & Medicine, 72, 1717-1727.

Saloniki, E.-C., Malley, J., Burge, P., Lu, H., Batchelder, L., Linnosmaa, I., et al. (2019). Comparing internet and face-to-face surveys as methods for eliciting preferences for social care-related quality of life: evidence from England using the ASCOT service user measure. Quality of Life Research, 1-14.

Swait, J., & Bernardino, A. (2000). Distinguishing taste variation from error structure in discrete choice data. Transportation Research Part B: Methodological, 34(1), 1-15.

Swait, J., & Louviere, J.J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of marketing research, 30(3), 305-314.

Towers, A. M., Smith, N., Palmer, S., Welch, E., & Netten, A. (2016). The acceptability and feasibility of using the Adult Social Care Outcomes Toolkit (ASCOT) to inform practice in care homes. BMC health services research, 16(1), 523.

Trukeschitz, B. (2011). "Worauf es letztlich ankommt. Ergebnisqualität in der Langzeitpflege und Betreuung." Kurswechsel 26(4): 22-35.

Ubel, P.A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Quality of life Research, 12(6), 599-607.

van Leeuwen, K. M., Bosmans, J. E., Jansen, A. P., Rand, S. E., Towers, A. M., Smith, N., ... & Ostelo, R. W. (2015). Dutch translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT). Health and quality of life outcomes, 13(1), 56.

Witt, J., Scott, A., & Osborne, R.H. (2009). Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. Health Economics, 18, 681-696.