Check for updates

*Original Article*

MDMP&P Policy & Practice

*MDM Policy & Practice*
1–16
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23814683211027902
journals.sagepub.com/home/mdm
SAGE

# Do You Prefer Safety to Social Participation? Finnish Population-Based Preference Weights for the Adult Social Care Outcomes Toolkit (ASCOT) for Service Users

**Lien Nguyen, Hanna Jokimäki, Ismo Linnosmaa, Eirini-Christina Saloniki, Laurie Batchelder, Juliette Malley, Hui Lu, Peter Burge, Birgit Trukeschitz, and Julien Forder**

**Abstract**

**Introduction.** The Adult Social Care Outcomes Toolkit (ASCOT) was developed in England to measure people's social care–related quality of life (SCRQoL). **Objectives.** The aim of this article is to estimate preference weights for the Finnish ASCOT for service users (ASCOT). In addition, we tested for learning and fatigue effects in the choice experiment used to elicit the preference weights. **Methods.** The analysis data ($n$ = 1000 individuals) were obtained from an online survey sample of the Finnish adult general population using gender, age, and region as quotas. The questionnaire included a best-worst scaling (BWS) experiment using ASCOT. Each respondent sequentially selected four alternatives (best, worst; second-best, second-worst) for eight BWS tasks ($n$ = 32,000 choice observations). A scale multinomial logit model was used to estimate the preference parameters and to test for fatigue and learning. **Results.** The most and least preferred attribute-levels were "I have as much control over my daily life as I want" and "I have no control over my daily life." The preference weights were not on a cardinal scale. The ordering effect was related to the second-best choices. Learning effect was in the last four tasks. **Conclusions.** This study has developed a set of preference weights for the ASCOT instrument in Finland, which can be used for investigating outcomes of social care interventions on adult populations. The learning effect calls for the development of study designs that reduce possible bias relating to preference uncertainty at the beginning of sequential BWS tasks. It also supports the adaptation of a modelling strategy in which the sequence of tasks is explicitly modelled as a scale factor.

**Keywords**

ASCOT, ASCOT for service users, best-worst scaling, Finland, learning and fatigue effects, preference, quality of life, scale multinomial logit, social care-related quality of life

Date received: November 2, 2020; accepted: June 1, 2021

## Introduction

The rising demand for long-term care (LTC) due to the ageing of the population raises the question of how public sector decision makers can effectively allocate limited resources within LTC systems to provide support for individuals with LTC needs and their informal carers.[1,2] Social care interventions support people in daily activities by enabling them to compensate for losses in

**Corresponding Author**
Lien Nguyen, Finnish Institute for Health and Welfare (THL), Mannerheimintie 166, PO Box 30, Helsinki, Uusimaa 00270, Finland; Telephone: (358) 29 524 7466 (lien.nguyen@thl.fi).

This Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

functional ability that are caused by a physical, mental, or emotional impairment or old age, with the final aim of enhancing well-being and quality of life (QoL).[3] Hence, the outcomes of social care interventions are broader than improvements in health.[4] Several systematic reviews have indicated that measurement of health alone, such as the quality-adjusted life year (QALY),[5] is insufficient to measure the effects of social care interventions.[6,7]

Population-based preferences have been used to assist priority-setting in many areas of public policy.[8,9] As a response to the need to have an outcome measure for social care interventions, Adult Social Care Outcome Toolkit (ASCOT) was developed in England.[10–12] The ASCOT draws on Sen's[13,14] capability approach to describe social care-related quality of life (SCRQoL) and is applicable to a range of different population groups and multiple care and support settings.[3] We focus on the ASCOT instrument for service users (hereafter ASCOT)—the four response-level interview and self-completion versions.

Internationally, there has been considerable interest in using ASCOT, for example, in Australia,[15] Austria,[16,17] Denmark,[18] Finland,[19] Italy,[20] Japan,[21] and the Netherlands.[22] At this time, the official Finnish translations of the ASCOT are available.[23] The Finnish ASCOT measure has been validated,[24] but its preference weights have not been estimated. Evidence indicates that country-specific differences in sociocultural values, demographic backgrounds, and political and economic systems influence valuations of well-being and QoL.[25,26] It has become a common practice in the field of health-related QoL measurement to establish country-specific preference weights to reflect country-specific values and perceptions concerning different health states.[27–29] A previous study showed that when comparing Spanish and UK time trade-off (TTO) values for EQ-5D health states, Spanish and UK values were similar for mild health states, but for health states worse than death, the Spanish weights generated lower utility scores than the UK weights.[30] Therefore, to enable the use of the Finnish ASCOT so that the ASCOT-QoL states of the Finnish population are correctly assessed, it is necessary to develop Finnish preference weights for the Finnish ASCOT measure.

Evidence from sequential choice experiments suggests that respondents' behavior may be influenced by fatigue (or boredom)* and learning.[31–33] Where respondents become fatigued, their engagement with the survey declines over time. This can be observed through greater inconsistency in the way they tend to respond to choice tasks toward the end of the choice experiment. Where respondents learn over the course of the choice tasks such that they become better at understanding the choice tasks, there is greater consistency in their responses toward the end of the choice experiment.[31–33] In practice, the model scale, capturing the variance of the error term, can allow for the effect of these behavioral mechanisms on estimated preference parameters.[34]

The primary aim of this study is to establish Finnish preference weights for the ASCOT, using data collected from a general population survey including a best-worst scaling (BWS) choice experiment.[35,36] As this experiment included a large number of sequential choices enabling the study of learning and fatigue, our secondary aim is to examine the existence of learning and fatigue effects in the BWS experiment. We will report here the conducted preference study and the estimation results, including evidence of the learning effect in the BWS experiment.

## Methods

### *The ASCOT and Best-Worst Scaling*

The ASCOT measures SCRQoL across eight attributes (domains): 1) control over daily life (CONT), 2) personal cleanliness and comfort (PERC), 3) food and drink (FOOD), 4) accommodation cleanliness and comfort (HOME), 5) personal safety (SAFE), 6) social participation and involvement (SOCI), 7) occupation (OCCU), and 8) dignity (DIGN). The basic domains are PERC, FOOD, HOME, and SAFE, and the rest are the higher order domains (Supplemental Table S1).[3] The dignity domain aims to capture the effects of the care process on the service user's self-esteem.[37] Attribute-levels indicating

Finnish Institute for Health and Welfare (THL), Helsinki, Finland (LN, HJ, IL); Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland (IL); Centre for Health Services Studies, University of Kent, Kent, UK (ECS); Personal Social Services Research Unit, University of Kent, Kent, UK (ECS, LB, JF); Care Policy and Evaluation Centre, London School of Economics and Political Science, London, UK (JM); RAND Europe, Cambridge, UK (HL, PB); Research Institute for Economics of Aging, WU Vienna University of Economics and Business, Vienna, Austria (BT). The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The EXCELC project was funded by the NORFACE Welfare State Futures programme under Grant Number 462-14-160. The Finnish contribution to the project was co-funded by the National Institute for Health and Welfare (THL), Finland, and the Austrian contribution by the Austrian Science Fund (FWF, Project Number I 2252-G16) and the Vienna Social Fund (FSW), Austria. The views expressed are not necessarily those of the funders.

*Although they are different in nature, boredom and fatigue similarly affect the results and one cannot distinguish between them.[34]

the intensity of needs are the following: ideal state (top level, level_1), no needs (level_2), some needs (level_3), and high needs (bottom level, level_4).

To collect preference data, we used the BWS method (profile case),[38] following Netten et al.[3] In a BWS choice experiment, respondents are asked to choose the most (best) and least preferred (worst) alternatives within each choice profile. In a traditional discrete choice experiment (DCE), respondents choose the most preferred alternative from at least two choice profiles at a time. In each BWS choice profile in this study, the same attributes were shown but their attribute-levels varied: respondents compared the displayed attribute-levels with each other to identify the best and worst, and the second-best and second-worst attribute-levels.[39,40] The BWS method is argued to provide easier tasks and to be less cognitively burdensome on the respondents compared to the DCE method.[35,39] It was also shown that using both DCE and BWS methods resulted in similar estimation results in the context of the ASCOT.[41]

## Design of Experiment and Survey, and Sampling

Different BWS scenarios were developed using an orthogonal main effects plan.[42] A fractional-factorial design was applied because a full-factorial design ($4^8$ possible profiles) would lead to far too many states for presentation.[43–45] We used a design matrix of 32 profiles which were divided into four 8-profile blocks (i.e., hypothetical scenarios describing the ASCOT attributes). A randomly chosen 8-profile block was shown to each respondent. The respondents were asked to do eight choice tasks by sequentially selecting the best, worst, second-best, and second-worst attribute-levels from each of eight BWS profiles (Supplemental Figure S1). A foldover design was applied to reduce the appearance of extreme values from each scenario (i.e., to eliminate easy or simple choices) and to ensure that both the occurrence of each attribute-level and the co-occurrences of the attribute-levels were equal.[46] The ordering of the attributes was also randomized between respondents (not within respondents) to prevent ordering bias and control for position effects on the selecting of attribute-levels.[47,48]

An online survey (managed by Research Now) was conducted between July and August 2016, using an internet panel as the sampling frame. To obtain a representative sample of the general adult population in Finland, a quota sampling approach was used with quotas for gender; age groups 18 to 24, 25 to 34, 35 to 44, 45 to 54, and ≥55 years old; and region. Regions in mainland Finland were included, and they were defined using the NUTS-2 classification (i.e., four large areas of Finland: West, Helsinki-Uusimaa, South, and East & North; Table 2). The questionnaire included questions about the respondents' demographic and socioeconomic background, well-being (self-assessed health [SAH] and overall QoL) and other information (the respondents' experience of caring and need for social care). The BWS section of the questionnaire also included questions about how well respondents understood the BWS tasks to provide insights into the validity and reliability of responses. To enhance the reliability of the results, respondents who spent less than 4.5 minutes completing the BWS tasks—an implausibly fast time based on piloting of the tasks—were excluded during the data collection phase. Excluding those with no information on their education ($n$ = 8), the analysis sample had 1000 respondents and the long-format panel data contained 32,000 choice observations (i.e., 32 choices per respondent).

## Modelling Strategy

Our choice data were collected using the BWS method,[42] which is a stated-preference method based in random utility theory (RUT).[36,54] Using RUT, one is able to elicit preferences for complex multidimensional commodities. RUT was extended to the case of discrete choices from multiple alternatives,[55] and further works have been developed (see Louviere et al.).[56] Similar to revealed preference approaches that measure and estimate preferences based on observations of individual choices, the BWS stated-preference method also assumes that respondents' choices reveal their utilities (preferences) and respondents choose [avoid] the alternative from which they will derive the highest utility [the lowest utility].[56] Researchers do not directly observe utility (i.e., utility is latent) but can observe choices made by respondents.[56] The expected utilities from choosing different alternatives were modelled in terms of attribute-levels rather than attributes of the individuals.[54,55] The preference parameters to be estimated are a function of choice frequencies, and the choice of a particular attribute-level describes the importance of that attribute-level relative to other attribute-levels.[42]

We applied models that described the choice process when the axioms of RUT do not fully hold. As a starting point, we used a traditional multinomial logit (MNL) model.[54,55] The basic MNL model is based on three assumptions: the independence of the errors; each error term follows a Gumbel distribution; and the errors are identically distributed[54,55] (see Appendix 1). Since the scale of the idiosyncratic error that captures the variance

**Table 1** Model Developing Process and Specifications

| Step | Model | Specification[a,b] | Result |
|---|---|---|---|
| 1. Basic model | MNL | Attribute position variables (separately for both the best or second-best and the worst or second-worst choices) as explanatory variables. | Model I |
| 2. Taste model[c] | Mixed logit | Including in the basic model (step 1): 1) the attribute-specific constants (ASCs) for the worst or second-worst choices, and 2) the interactions between the individual characteristics (e.g., age, gender, education) and the attribute-levels or single attributes to measure taste heterogeneity. We aimed to control for taste heterogeneity and to minimize unexplained variations. | Model II: Not reported |
| 3. Taste-and-scale model | G-MNL | Including in the taste model (step 2): different sets of 4 to 5 variables at a time to test whether these variables could capture scale heterogeneity. | Model III: Supplemental Table S2 |
| 4. Scale model | S-MNL | Keeping the scale variables obtained from step 3 and the position variables. The ASCs for the worst or second-worst choices and the variables capturing taste heterogeneity were excluded. | Model IV |
| 5. Taste-adjusted scale model | S-MNL with taste variables | Including in the scale model (step 4) several taste variables. These were interaction terms between attribute-levels and observed characteristics of respondents that were not representative of the Finnish general adult population. | Model V: Supplemental Table S3 |
| | | Final preference weights | Model V+ |

G-MNL, generalized multinomial logit; MNL, multinomial logit; S-MNL, scale multinomial logit.
[a]Each model always included attribute-level variables that we were interested in.
[b]The specified variables were included in the model as explanatory variables.
[c]The interaction terms (the taste variables) measured the impacts that individual characteristics had on the preferences for a particular attribute-level or attribute.

of the error term in the MNL model is usually normalized to a constant (generally to unity) to ensure identification of the model parameters, scale heterogeneity is often ignored.[57] To use the traditional MNL model, the assumption of independence of irrelevant alternatives should be valid, but proportional substitution across alternatives is likely to happen in the actual data.[58]

Scale heterogeneity can distort preference parameters.[59] Hence, we considered the scale multinomial logit (S-MNL) model to be more suitable than the basic MNL model to derive preference parameters.[57] To investigate which factors scale heterogeneity in the sample was associated with, we focused on the heterogeneity in error variance that was not accounted for by taste heterogeneity related to observed characteristics of respondents. Therefore, before selecting the final S-MNL model for the estimation of preference parameters, we first ran the mixed logit model using observed characteristics of respondents (e.g., age, gender, and education) to examine taste heterogeneity only, and then ran the generalized MNL (G-MNL) model to examine scale heterogeneity, simultaneously controlling for taste heterogeneity.[57] The G-MNL model enabled us to find appropriate scale factors. These models are defined in more detail in Appendix 1.[55,57,58,60–62]

Our five-step modelling process is described in Table 1. For simplicity's sake, we have named the five specifications of the G-MNL as follows: 1) basic (Model I), 2) taste (Model II), 3) taste-and-scale (Model III), 4) scale (Model IV), and 5) taste-adjusted scale (Model V) models (Table 1). The estimated attribute-level coefficients from Model V were then adjusted for significant taste differences between the sample population and the general

**Table 2** Descriptive Characteristics of the Sample Population and the General Population

| Variable | Sample ($n$ = 1000) Mean | SD | General Population Mean | $N$ | Source |
|---|---|---|---|---|---|
| Gender | | | 1.000 | 4,431,392 | Statistics Finland[49] |
| Male | 0.481 | 0.500 | 0.488 | 2,163,845 | |
| Female | 0.519 | 0.500 | 0.512 | 2,267,547 | |
| Age | | | 1.000 | 4,431,392 | Statistics Finland[49] |
| 18–24 years | 0.078 | 0.268 | 0.103 | 455,977 | |
| 25–34 years | 0.155 | 0.362 | 0.159 | 704,402 | |
| 35–44 years | 0.160 | 0.367 | 0.151 | 671,350 | |
| 45–54 years | 0.185 | 0.388 | 0.161 | 712,553 | |
| 55–64 years | 0.272 | 0.445 | 0.166 | 737,135 | |
| 65–79 years | 0.145 | 0.352 | 0.194 | 861,876 | |
| 80 years or older | 0.005 | 0.071 | 0.065 | 288,099 | |
| Marital status | | | 1.000 | 4,431,392 | Statistics Finland[49] |
| Married | 0.417 | 0.493 | 0.451 | 1,998,678 | |
| Divorced | 0.188 | 0.391 | 0.128 | 568,184 | |
| Widowed | 0.027 | 0.162 | 0.064 | 282,794 | |
| Single | 0.346 | 0.476 | 0.357 | 1,581,736 | |
| Not reported | 0.022 | 0.147 | | | |
| Employment status | | | 1.000 | 4,431,392 | Statistics Finland[50] |
| Employed | 0.429 | 0.495 | 0.514 | 2,275,679 | |
| Student | 0.064 | 0.245 | 0.053 | 236,335 | |
| Pensioner | 0.296 | 0.456 | 0.314 | 1,389,266 | |
| Unemployed | 0.143 | 0.350 | 0.080 | 355,364 | |
| Other | 0.068 | 0.252 | 0.039 | 174,748 | |
| Education | | | 1.000 | 4,591,285 | Statistics Finland[51,a] |
| Lower secondary school | 0.092 | 0.289 | 0.293 | 1,345,561 | |
| Upper secondary school | 0.484 | 0.500 | 0.407 | 1,867,828 | |
| Lowest level tertiary school | 0.126 | 0.332 | 0.097 | 447,112 | |
| Lower level tertiary school | 0.175 | 0.380 | 0.105 | 484,271 | |
| Higher level tertiary school | 0.114 | 0.318 | 0.088 | 403,731 | |
| Doctorate level | 0.009 | 0.094 | 0.009 | 42,782 | |
| Housing tenure | | | 1.000 | 5,363,637 | Statistics Finland[52,a] |
| Own house/apartment | 0.546 | 0.498 | 0.711 | 3,813,335 | |
| Rent | 0.447 | 0.497 | 0.270 | 1,446,729 | |
| Other | 0.007 | 0.083 | 0.019 | 103,573 | |
| Religion | | | 1.000 | 4,609,119 | Statistics Finland[53,a] |
| Any religion | 0.620 | 0.485 | 0.733 | 3,376,789 | |
| No religion | 0.380 | 0.485 | 0.267 | 1,232,330 | |
| Regions | | | 1.000 | 4,407,913 | Statistics Finland[49,a] |
| Helsinki and Uusimaa | 0.291 | 0.454 | 0.297 | 1,311,203 | |
| Southern Finland | 0.219 | 0.414 | 0.215 | 948,790 | |
| Western Finland | 0.251 | 0.434 | 0.252 | 1,110,490 | |
| North-Eastern Finland | 0.239 | 0.426 | 0.235 | 1,037,430 | |

[a]Religion (Statistics Finland)[49] and education (Statistics Finland)[51] refer to the population aged 15 or older. Housing tenure (Statistics Finland)[52] refers to the whole housing population and regions (Statistics Finland)[49] to the population aged 18 or older.

population using modified post-stratification to become final preference weights (Model V+).[63]

In all models, in addition to attribute-levels, the position variables for the best or second-best choices and those for the worst or second-worst choices were included to account for the overall effects of attribute ordering associated with the experimental choice task design.[64,65] Furthermore, we included the attribute-specific constants (ASCs) for the worst or second-worst choices in the taste and taste-and-scale models (Supplemental Table S2) to capture differences in the likelihoods of selecting attributes as the least or second-least preferred alternatives. The ASC is assumed to account for the average effect on the utility of all factors that are not included in the model.[66]

The bottom level of the control attribute, *cont4*, was used as the reference level and was set to zero. To ensure model identification, the position variable of one attribute (for each set of best or worst choices) were set at zero, and the ASC of one attribute in the scenario for the worst or second-worst choices in the taste and taste-and-scale MNL models was fixed at zero. Due to the repeated choice data, sandwich estimators were used to get robust standard errors.[66] The variables capturing taste heterogeneity were obtained using the "Apply Run" procedure in ALOGIT.[67] The models were estimated by the maximum likelihood using the BIOGEME software.[68]

### *Scale Heterogeneity and Behavioral Mechanisms*

We tested whether scale factors were associated with age, gender, education, residential area, housing tenure, SAH, overall QoL, experience of care, time to complete the BWS tasks, and the best and worst choices. To select scale variables, we undertook a series of scale heterogeneity analyses—each time we estimated a taste-and-scale model by including in it four to five covariates above with different subgroups of each covariate.

Fatigue and learning may arise from the repeated and sequential choice tasks in choice experiments, which can influence the respondents' preferences.[31–33] Hence, we expected that a scale factor may be associated with the sequence of the BWS choice tasks. Following Carlsson et al.,[32] we defined two identical sequences of four tasks in the BWS experiment. We tested for the presence of learning [or fatigue] in the first sequence of four tasks relative to the second sequence of four tasks. Using this specification, learning [fatigue] suggests that the error variance is higher [lower] in the first four-task sequence than in the second four-task sequence.[32,34]

### *Final Preference Estimates*

The established preference weights should be representative of the general population, being the averages of the preference estimates for all individuals.[42] Several subgroups of covariates in the sample were significantly under- or overrepresented when comparing them to the general population (>10 percentage points of $P < 0.05$). This was the case for those aged 65 or older, those aged 54 to 64, and those without any religion (Table 2). For these subgroups, applying a post-stratification method,[63] we used group-specific population weights to adjust the preference estimates from the taste-adjusted S-MNL model* (Model V; Supplemental Table S3).[3,69,70] The final preference estimates can then reflect the general population's values (Model V+; Table 4). We also computed the standard errors of the adjusted preference estimates, using population weights (Table 2) and the estimated variance-covariance matrix of the parameters provided by BIOGEME.[68]

We normalized the attribute-level coefficients, using the largest coefficient from each estimated model as the common denominator. To better understand the quantified changes in different SCRQoL states, we linearly rescaled the final attribute-level coefficients to an index by applying a conversion method.[44,69,70] We anchored this index at a value of zero for the set of states presented by the eight lowest attribute-level coefficients (each per domain) and at a value of one for the set of states presented by the eight highest attribute-level coefficients (each per domain), keeping the relative differences between the attribute-level coefficients unchanged. Thus, the ASCOT index measuring SCRQoL ranges between zero and one, where zero indicates the worst SCRQoL represented by the eight worst ASCOT-QoL states and one indicates the best SCRQoL represented by the eight best ASCOT-QoL states.

## Results

### *Sample Characteristics*

The sample proportions of many covariates were close to those of the Finnish general adult population. We found larger differences between the sample and the general population for age, education, religion, and housing tenure than for the other variables (Table 2). Compared to the general population, the sample had more people aged 55 to 64 years and fewer those aged 80 or older, fewer people with the lowest level of education and belonging to some religion, as well as fewer house owners.

Regarding how often respondents were able to put themselves in the imaginary situations in the BWS exercises, 67% of them were able to do so all of the time and 29.6% some of the time. Nearly everyone reported that they had understood the situations in the best-worst exercises all or some of the time (99.7%; Table 3).

The three attributes mostly preferred were the control (CONT), food (FOOD), and occupation (OCCU)

*The taste-adjusted S-MNL model included the attribute-levels, the position variables for the best or second-best and worst or second-worst choices, the identified scale factors, and the identified interaction terms between attribute-levels and unrepresentative subgroups of covariates, but did not include the ASCs for the worst or second-worst choices (Model V; Supplemental Table S).

**Table 3** Descriptive Statistics of Variables Used in the Analysis ($n = 32{,}000$)[a]

| **Description**[b] | **Descriptive Value** | | | |
|---|---|---|---|---|
| | **Name** | **Mean** | | |
| | | **All** | **Best/Second-Best Choice** | **Worst/Second-Worst Choice** |
| *Attributes or attribute-levels* | | | | |
| Control over daily life | CONT | 0.173 | 0.217 | 0.129 |
| 1 I have as much control over my daily life as I want | cont1 | 0.048 | 0.091 | 0.004 |
| 2 I have adequate control over my daily life | cont2 | 0.048 | 0.091 | 0.005 |
| 3 I have some control over my daily life, but not enough | cont3 | 0.030 | 0.031 | 0.028 |
| 4 I have no control over my daily life | cont4 | 0.048 | 0.004 | 0.091 |
| Personal cleanliness and comfort | PERC | 0.103 | 0.091 | 0.116 |
| 1 I feel clean and am able to present myself the way I like | perc1 | 0.024 | 0.044 | 0.005 |
| 2 I feel adequately clean and presentable | perc2 | 0.022 | 0.040 | 0.005 |
| 3 I feel less than adequately clean or presentable | perc3 | 0.030 | 0.004 | 0.056 |
| 4 I don't feel at all clean or presentable | perc4 | 0.027 | 0.003 | 0.051 |
| Food and drink | FOOD | 0.159 | 0.145 | 0.173 |
| 1 I get all the food and drink I like when I want | food1 | 0.035 | 0.066 | 0.004 |
| 2 I get adequate food and drink at OK times | food2 | 0.037 | 0.068 | 0.005 |
| 3 I don't always get adequate or timely food and drink | food3 | 0.041 | 0.006 | 0.075 |
| 4 I don't always get adequate or timely food and drink, and I think there is a risk to my health | food4 | 0.047 | 0.005 | 0.088 |
| Accommodation cleanliness and comfort | HOME | 0.061 | 0.063 | 0.059 |
| 1 My home is as clean and comfortable as I want | home1 | 0.020 | 0.035 | 0.006 |
| 2 My home is adequately clean and comfortable | home2 | 0.013 | 0.022 | 0.005 |
| 3 My home is not quite clean or comfortable enough | home3 | 0.011 | 0.003 | 0.019 |
| 4 My home is not at all clean or comfortable | home4 | 0.017 | 0.003 | 0.030 |
| Personal safety | SAFE | 0.120 | 0.067 | 0.173 |
| 1 I feel as safe as I want | safe1 | 0.026 | 0.047 | 0.004 |
| 2 Generally I feel adequately safe, but not as safe as I would like | safe2 | 0.021 | 0.014 | 0.028 |
| 3 I feel less than adequately safe | safe3 | 0.033 | 0.003 | 0.062 |
| 4 I don't feel at all safe | safe4 | 0.041 | 0.003 | 0.078 |
| Social participation and involvement | SOCI | 0.098 | 0.109 | 0.086 |
| 1 I have as much social contact as I want with people I like | soci1 | 0.031 | 0.057 | 0.005 |
| 2 I have adequate social contact with people | soci2 | 0.022 | 0.041 | 0.004 |
| 3 I have some social contact with people, but not enough | soci3 | 0.013 | 0.009 | 0.017 |
| 4 I have little social contact with people and feel socially isolated | soci4 | 0.032 | 0.003 | 0.061 |
| Occupation | OCCU | 0.157 | 0.207 | 0.107 |
| 1 I'm able to spend my time as I want, doing things I value or enjoy | occu1 | 0.044 | 0.086 | 0.003 |
| 2 I'm able do enough of the things I value or enjoy with my time | occu2 | 0.045 | 0.085 | 0.004 |
| 3 I do some of the things I value or enjoy with my time, but not enough | occu3 | 0.030 | 0.032 | 0.028 |
| 4 I don't do anything I value or enjoy with my time | occu4 | 0.038 | 0.004 | 0.072 |
| Dignity | DIGN | 0.128 | 0.100 | 0.156 |
| 1 The way I'm helped and treated makes me think and feel better about myself | dign1 | 0.034 | 0.062 | 0.006 |
| 2 The way I'm helped and treated does not affect the way I think or feel about myself | dign2 | 0.019 | 0.028 | 0.010 |
| 3 The way I'm helped and treated sometimes undermines the way I think and feel about myself | dign3 | 0.031 | 0.007 | 0.056 |
| 4 The way I'm helped and treated completely undermines the way I think and feel about myself | dign4 | 0.044 | 0.004 | 0.084 |

(*continued*)

**Table 3** (continued)

| *Other variables* | Name | Mean | SD |
|---|---|---|---|
| Attribute position | | | |
| For best/second-best choices | | | |
| Attribute appeared in the 1st row | pos1_B | 0.071 | 0.257 |
| Attribute appeared in the 2nd row | pos2_B | 0.067 | 0.251 |
| Attribute appeared in the 3rd row | pos3_B | 0.064 | 0.244 |
| Attribute appeared in the 4th row | pos4_B | 0.064 | 0.244 |
| Attribute appeared in the 5th row | pos5_B | 0.060 | 0.237 |
| Attribute appeared in the 6th row | pos6_B | 0.060 | 0.237 |
| Attribute appeared in the 7th row | pos7_B | 0.056 | 0.230 |
| Attribute appeared in the 8th row | pos8_B | 0.059 | 0.235 |
| For worst/second-worst choices | | | |
| Attribute appeared in the 1st row | pos1_W | 0.060 | 0.238 |
| Attribute appeared in the 2nd row | pos2_W | 0.063 | 0.243 |
| Attribute appeared in the 3rd row | pos3_W | 0.062 | 0.242 |
| Attribute appeared in the 4th row | pos4_W | 0.064 | 0.245 |
| Attribute appeared in the 5th row | pos5_W | 0.062 | 0.241 |
| Attribute appeared in the 6th row | pos6_W | 0.064 | 0.244 |
| Attribute appeared in the 7th row | pos7_W | 0.063 | 0.243 |
| Attribute appeared in the 8th row | pos8_W | 0.062 | 0.241 |
| Scale consistency influenced[c] by | | | |
| Learning (the first four BWS tasks) | learning | 0.500 | 0.500 |
| Self-assessed health (SAH) as fair or bad or very bad | sah | 0.447 | 0.497 |
| An upper-secondary school or lower level | edu | 0.702 | 0.457 |
| Short completion time (= the first quartile of the distribution of time used to complete the BWS tasks; 7.27 minutes at the most) | time | 0.252 | 0.434 |
| Understanding the tasks | | | |
| (1) Did you feel that you could put yourself in the imaginary situations described in the best-worst exercises? | | | |
| Yes, all of the time | | 0.670 | |
| Yes, but only some of the time | | 0.296 | |
| No | | 0.034 | |
| (2) In the best-worst exercises, did you understand the situations? | | | |
| Yes, all of them | | 0.869 | |
| Yes, but only some of them | | 0.128 | |
| No | | 0.003 | |

[a]The first column introduces 8 attributes, 32 attribute-levels (4 per attribute), 16 attribute position variables (8 for best/second-best choices; 8 for worst/second-worst choices), and four variables capturing scale heterogeneity. The second column indicates empirical names for the variables used in the models. For the attributes and attribute-levels, the next three columns describe the proportions of the attributes or attribute-levels that respondents chose totally and by choice type. For the attribute position variables, the third and fourth columns describe the sample mean and standard deviation of each attribute position variable. For two questions that explained how respondents understood the BWS tasks in the experiment, the proportions of multiple response items are reported.
[b]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 finascot@thl.fi to enquire about permissions.
[c]The reference group 1) for learning: the last four BWS tasks; 2) for SAH: very good or good SAH; 3) for education: lowest or lower or higher level tertiary school or doctorate level; 4) for short completion time: longer completion time (= the second or third or fourth quartile of the distribution of time used to complete the BWS tasks).

attributes (Table 2). The *cont1*, *cont2*, *occu1*, and *occu2* attribute-levels were most preferred: they were mostly selected as the best or second-best choices. The *cont4*, *food4*, *dign4*, and *safe4* attribute-levels were least preferred: they were mostly chosen as the worst or second-worst choices. In particular, *food2* was preferred to *food1*; it was selected more often than *food1* in total and across choices. For the best or second-best choices, until the seventh position, the further away from the first attribute on the choice list, the less likely it was that an attribute was chosen. For the worst or second-worst choices, the position of the attribute did not matter a great deal: the probability of choosing an attribute remained quite stable from the first to the eight position.

## *Preference Estimates*

The final preference estimates (Model V+; Table 4) were derived using results from the taste-adjusted scale MNL (Model V; Supplemental Table S3).* Concerning the goodness-of-fit, Model IV was significantly better than Model I, and Model V+ was significantly better than Model IV.† Because a pseudo-$R^2$ with values between the range of 0.3 and 0.4 can be regarded as an $R^2$ with values between the range of 0.6 and 0.8 for the equivalent linear regression,[71] the pseudo-$R^2$ of 0.298 (Models IV and $V^+$) suggests a decent fit. Below, we discuss the estimation results from Model V+ if not otherwise specified.

The *cont4* attribute-level, "I have no control over my daily life," the lowest-valued state, was followed by the attribute-levels *food4*, *dign4*, *food3*, and *occu4* in an ascending order (Table 4; Figure 1). Since the coefficients of the other attribute-levels were greater than zero, the other ASCOT-QoL states were valued more than the ASCOT-QoL state represented by *cont4*. The *cont1* attribute-level, the mostly valued state, was followed by *cont2* and the top two levels of the OCCU attribute in a descending order.

The ordering of the attribute-level coefficients by each attribute followed the original ordering of the attribute-levels except for the FOOD attribute. The difference between attribute-levels 1–2 was not statistically significant for attributes: CONT ($P$ = 0.555), PERC ($P$ = 0.375), FOOD ($P$ = 0.864, with the difference in switched attribute-levels 2–1), HOME ($P$ = 0.228), and OCCU ($P$ = 0.760). The ASCOT-QoL states indicated by attribute-levels 1–2 were valued more than those indicated by attribute-levels 3–4. The change in ASCOT-QoL due to moving from level_3 to level_2 was valued more than that due to the moving from level_2 to level_1. However, for the SAFE and SOCI attributes, the mostly valued change in ASCOT-QoL was associated with the movement from *safe2* [no needs] to *safe1* [ideal state] and that from *soci4* [high needs] to *soci3* [some needs] (Table 4; Figure 1).

For the best or second-best choices, the coefficients of the position variables were statistically significant (Table 4). As implied by the negative sign of the coefficients, respondents were less likely to choose an item on the choice list that appeared following the first item. Furthermore, the attributes were increasingly less likely to be chosen the further down the list they appeared, until reaching the seventh position (–0.503), which was less likely to be chosen than the eight position (–0.402). However, respondents were quite indifferent to the items appearing on the fifth to sixth rows. The coefficients of the position variables for the worst or second-worst choices were not statistically significant.

The value of each scale parameter was smaller than unity.‡ Respondents who had a high level of education or a better SAH, or took a longer time doing the BWS tasks were more consistent in their choices than those who had a lower level of education or worse SAH, or took less time doing the BWS tasks (Table 4). Moreover, the estimate of the scale parameter for learning (0.889) indicates that the error variance was higher in the first four-task sequence than in the second four-task sequence. This evidence shows that learning occurred in the later four tasks of the sequential BWS choice experiment.

## *Final Preference Weights*

Table 4 also reports normalized and rescaled values of the attribute-level coefficients. Due to disparities between each attribute-level and the average value of all lowest rated attribute-levels,[44,70] their rescaled values were also negative (Tables 4 and 5). Regarding the FOOD attribute, *food1* had a smaller coefficient (5.845) than *food2* (5.888), although their difference was not statistically significant

*Supplemental Table S2 reports results from the taste-and-scale MNL model (Model III).

†Regarding Models I and IV, the LR test statistic was $-2 \times \{-42056.3 - (-41711.9)\} = 688.8$, with df = 49 – 45 = 4, and $P < 0.001$, which was in favor of Model IV. A correspondent LR test statistic for Models IV and $V^+$ was $-2 \times \{-41711.9 - (-41698.4)\} = 27.0$, with df = 54 – 49 = 5, and $P < 0.001$, which was in favor of Model V+. $Rho^2$ produced by BIOGEME[68] is pseudo-$R^2$.

‡The value of each scale parameter ($\lambda$) is inversely related to the level of the error variance of the tested group compared to the reference group. If $\lambda$ is smaller (greater) than unity, the tested group has higher (lower) error variance compared to the reference group.

**Table 4** Estimated Finnish Preference Weights for the ASCOT for Service Users ($n = 32{,}000$)

| | Model I | | | Model IV | | | Model V+ [a,b] | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Estimated coeff.** | **Robust *t*-value** | **Normalized coeff.** | **Estimated coeff.** | **Robust *t*-value** | **Normalized coeff.** | **Estimated coeff.** | **Robust *t*-value** | **Normalized coeff.** | **Rescaled coeff.** |
| *Attribute-level* | | | | | | | | | | |
| cont1 | 5.077 | 41.96 | 1.000 | 7.086 | 23.27 | 1.000 | 6.903 | 23.44 | 1.000 | 0.156 |
| cont2 | 4.986 | 41.48 | 0.982 | 6.917 | 23.24 | 0.976 | 6.731 | 23.39 | 0.975 | 0.152 |
| cont3 | 2.524 | 30.68 | 0.497 | 3.436 | 20.34 | 0.485 | 3.262 | 20.45 | 0.473 | 0.063 |
| cont4 | 0.000 | — | 0.000 | 0.000 | — | 0.000 | 0.000 | — | 0.000 | –0.020 |
| perc1 | 3.734 | 38.35 | 0.735 | 5.198 | 22.62 | 0.734 | 5.017 | 22.73 | 0.727 | 0.108 |
| perc2 | 3.589 | 36.68 | 0.707 | 5.003 | 22.13 | 0.706 | 4.823 | 22.26 | 0.699 | 0.103 |
| perc3 | 1.239 | 20.13 | 0.244 | 1.673 | 16.29 | 0.236 | 1.510 | 16.21 | 0.219 | 0.018 |
| perc4 | 1.195 | 20.70 | 0.235 | 1.584 | 16.89 | 0.224 | 1.420 | 16.75 | 0.206 | 0.016 |
| food1 | 4.353 | 41.18 | 0.857 | 6.048 | 23.07 | 0.854 | 5.845 | 23.11 | 0.847 | 0.129 |
| food2 | 4.386 | 41.31 | 0.864 | 6.091 | 23.66 | 0.860 | 5.888 | 23.74 | 0.853 | 0.130 |
| food3 | 0.627 | 11.47 | 0.124 | 0.845 | 10.52 | 0.119 | 0.607 | 6.94 | 0.088 | –0.005 |
| food4 | 0.243 | 4.19 | 0.048 | 0.311 | 3.91 | 0.044 | 0.055 | 0.63 | 0.008 | –0.019 |
| home1 | 3.378 | 37.55 | 0.665 | 4.679 | 22.10 | 0.660 | 4.478 | 22.30 | 0.649 | 0.094 |
| home2 | 3.168 | 34.35 | 0.624 | 4.418 | 21.60 | 0.623 | 4.239 | 21.72 | 0.614 | 0.088 |
| home3 | 2.097 | 30.68 | 0.413 | 2.835 | 21.36 | 0.400 | 2.670 | 21.70 | 0.387 | 0.048 |
| home4 | 1.694 | 28.89 | 0.334 | 2.271 | 20.31 | 0.320 | 2.109 | 20.74 | 0.306 | 0.034 |
| safe1 | 3.923 | 37.65 | 0.773 | 5.497 | 22.31 | 0.776 | 5.325 | 22.34 | 0.771 | 0.116 |
| safe2 | 2.284 | 31.53 | 0.450 | 3.128 | 21.06 | 0.441 | 2.893 | 20.95 | 0.419 | 0.054 |
| safe3 | 1.066 | 17.83 | 0.210 | 1.397 | 14.70 | 0.197 | 1.095 | 13.11 | 0.159 | 0.008 |
| safe4 | 0.607 | 10.91 | 0.120 | 0.787 | 9.77 | 0.111 | 0.669 | 6.87 | 0.097 | –0.003 |
| soci1 | 4.078 | 37.68 | 0.803 | 5.716 | 21.93 | 0.807 | 5.535 | 22.03 | 0.802 | 0.121 |
| soci2 | 3.656 | 37.62 | 0.720 | 5.090 | 22.26 | 0.718 | 4.908 | 22.42 | 0.711 | 0.105 |
| soci3 | 2.334 | 32.53 | 0.460 | 3.213 | 21.30 | 0.453 | 3.047 | 21.56 | 0.441 | 0.058 |
| soci4 | 0.912 | 16.30 | 0.180 | 1.186 | 13.98 | 0.167 | 1.031 | 13.27 | 0.149 | 0.006 |
| occu1 | 4.859 | 41.25 | 0.957 | 6.763 | 23.09 | 0.954 | 6.544 | 23.16 | 0.948 | 0.147 |
| occu2 | 4.785 | 41.26 | 0.942 | 6.644 | 23.32 | 0.938 | 6.459 | 23.49 | 0.936 | 0.145 |
| occu3 | 2.549 | 32.73 | 0.502 | 3.490 | 21.46 | 0.493 | 3.315 | 21.74 | 0.480 | 0.065 |
| occu4 | 0.634 | 12.26 | 0.125 | 0.814 | 10.68 | 0.115 | 0.653 | 9.23 | 0.095 | –0.003 |
| dign1 | 4.301 | 38.65 | 0.847 | 5.975 | 22.57 | 0.843 | 5.792 | 22.67 | 0.839 | 0.128 |
| dign2 | 3.212 | 31.72 | 0.633 | 4.475 | 20.87 | 0.632 | 4.297 | 21.03 | 0.622 | 0.090 |
| dign3 | 1.225 | 20.21 | 0.241 | 1.606 | 16.29 | 0.227 | 1.444 | 15.89 | 0.209 | 0.017 |
| dign4 | 0.406 | 7.21 | 0.080 | 0.516 | 6.75 | 0.073 | 0.356 | 5.00 | 0.052 | –0.011 |
| *Position for best/second-best (most/second most preferred) alternatives* | | | | | | | | | | |
| pos2_B | –0.110 | –2.28 | | –0.148 | –2.25 | | –0.147 | –2.23 | | |
| pos3_B | –0.192 | –3.96 | | –0.255 | –3.74 | | –0.255 | –3.75 | | |
| pos4_B | –0.204 | –4.15 | | –0.245 | –3.62 | | –0.246 | –3.63 | | |
| pos5_B | –0.308 | –5.84 | | –0.404 | –5.55 | | –0.401 | –5.52 | | |
| pos6_B | –0.319 | –6.24 | | –0.405 | –5.70 | | –0.407 | –5.74 | | |
| pos7_B | –0.395 | –7.64 | | –0.501 | –6.99 | | –0.503 | –7.04 | | |
| pos8_B | –0.323 | –6.17 | | –0.401 | –5.59 | | –0.402 | –5.62 | | |
| *Position for worst/second-worst (least/second least preferred) alternatives* | | | | | | | | | | |
| pos2_W | 0.027 | 0.55 | | 0.040 | 0.60 | | 0.044 | 0.66 | | |
| pos3_W | 0.037 | 0.74 | | 0.047 | 0.68 | | 0.052 | 0.75 | | |
| pos4_W | 0.015 | 0.31 | | 0.023 | 0.35 | | 0.027 | 0.41 | | |
| pos5_W | 0.073 | 1.44 | | 0.090 | 1.31 | | 0.097 | 1.43 | | |
| pos6_W | 0.036 | 0.74 | | 0.051 | 0.77 | | 0.058 | 0.87 | | |
| pos7_W | 0.057 | 1.13 | | 0.070 | 1.00 | | 0.076 | 1.09 | | |
| pos8_W | 0.056 | 1.09 | | 0.070 | 0.99 | | 0.066 | 0.94 | | |
| *Scale parameter* | | | | | | | | | | |
| $\lambda_{learning}$ | | | | 0.893 | 6.46 | | 0.889 | 6.71 | | |
| $\lambda_{sah}$ | | | | 0.811 | 5.34 | | 0.812 | 5.29 | | |
| $\lambda_{edu}$ | | | | 0.861 | 3.56 | | 0.864 | 3.46 | | |
| $\lambda_{time}$ | | | | 0.748 | 6.01 | | 0.748 | 6.04 | | |

*(continued)*

**Table 4** (continued)

| Variable | Model I | | | Model IV | | | Model V+ [a,b] | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Estimated coeff. | Robust *t*-value | Normalized coeff. | Estimated coeff. | Robust *t*-value | Normalized coeff. | Estimated coeff. | Robust *t*-value | Normalized coeff. | Rescaled coeff. |
| Observations | | 32,000 | | 32,000 | | | 32,000 | | | |
| df | | 45 | | 49 | | | 54 | | | |
| Log likelihood value | | –42056.3 | | –41711.9 | | | –41698.4 | | | |
| AIC | | 84202.6 | | 83521.8 | | | 83504.8 | | | |
| BIC | | 84315.3 | | 83644.6 | | | 83640.1 | | | |
| $Rho^2$ (0) | | 0.292 | | 0.298 | | | 0.298 | | | |

AIC, Akaike information criterion; BIC, Bayesian information criterion.
[a]The coefficients of the attribute-levels in Model V+ were adjusted for the observed taste differences between the sample and general populations. Corrections were manually made for 10 attribute-levels (occu1, home1, and all four levels of both FOOD and SAFE attributes).
[b]Five interaction terms capturing taste heterogeneity were included in the taste-adjusted S-MNL model (df = 54) (Supplemental Table S3).

(Table 4). To preserve the rank of the attribute-levels and better quantify changes in SCRQoL in practice, we switched these estimated coefficients. Therefore, the normalized (rescaled) preference weight of *food1* was 0.853 (0.130) and that of *food2* was 0.847 (0.129) (Table 5).

If we do not need preference weights for four separate attribute-levels, we can combine the parameters of *food1* and *food2*. Using the restriction that the parameters of *food1* and *food2* are same, we ran a new taste-adjusted S-MNL model (log-likelihood value –41698.55; df = 53; $Rho^2 = 0.298$). Regarding the goodness-of-fit, this model was supported by the LR test (statistic 0.30; df = 1) compared to Model V+ (Table 4). The estimated coefficient of the combined attribute-level (called *food12*) was exactly the average of the original coefficients of *food1* and *food2*, and the rest of the estimated parameters were very similar to those in Model V+. Nevertheless, to preserve the order of the attribute-levels indicating the intensity of need for each ASCOT attribute, we switched the estimated preference weights of the top two levels of FOOD (Table 5).

ASCOT index values can be used to illustrate changes in SCRQoL associated with different ASCOT-QoL states. As individuals' SCRQoL is an additive combination of eight attribute-levels, an improvement in SCRQoL, for example, from a poor state of 34424343[*] to an improved state of 12313212 would be a SCRQoL gain of 0.481[†] (Supplemental Figure S2). Those who would like to utilize the Finnish ASCOT can use the normalized or rescaled values of the preference weights displayed in Table 5.

## Discussion

In this study, we derived the Finnish population-based preference weights for the Finnish ASCOT and provided evidence on the learning effect in the BWS choice experiment. Although population-based preferences cannot fully capture population-specific expectations and responses,[72] they are regarded as appropriate for evaluating the effects of social care interventions for adult populations.[3]

The Finnish respondents placed both highest and lowest values to attribute-levels of the higher-order control and occupation attributes. A comparison of the Finnish preference weights for the ASCOT instrument to the English,[3] Austrian,[17] and Japanese weights[73] also suggests that the control and occupation attributes were mostly valued (Supplemental Tables S4). Although the estimated models had different scale factors,[74] we can study the relative size of the differences in SCRQoL by using one of the attribute-levels as a common denominator.[3] The most valued attribute-level was the *cont1* state in Finland and England, while it was the *occu1* state in Austria and Japan. The least valued attribute-level was the *cont4* state in Finland and England, but it was the *dign4* state in Austria and Japan. In fact, *cont4* and *dign4* were equally the lowest valued states in Japan. The Finnish and English preference weights were very highly positively correlated—the Pearson correlation coefficient was 0.97 ($P < 0.0001$)—indicating strong consistency between the weights.

We found evidence that the probability of an item being selected depended on its position in the profile but

*The attributes were specified in the following order: 1) CONT, 2) PERC, 3) FOOD, 4) HOME, 5) SAFE, 6) SOCI, 7) OCCU, and 8) DIGN (Table 3). Thus, the state of 34424343 consisted of cont3, perc4, food4, home2, safe4, soci3, occu4, and dign3 attribute-levels and that of 12313212 consisted of cont1, perc2, food3, home1, safe3, soci2, occu1, and dign2 attribute-levels.

†There was a change in value from 0.217 [= 0.063 + 0.016 + (–0.019) + 0.088 + (–0.003) + 0.058 + (–0.003) + 0.0168] to 0.698 [= 0.156 + 0.103 + (–0.005) + 0.094 + 0.008 + 0.105 + 0.147 + 0.09].

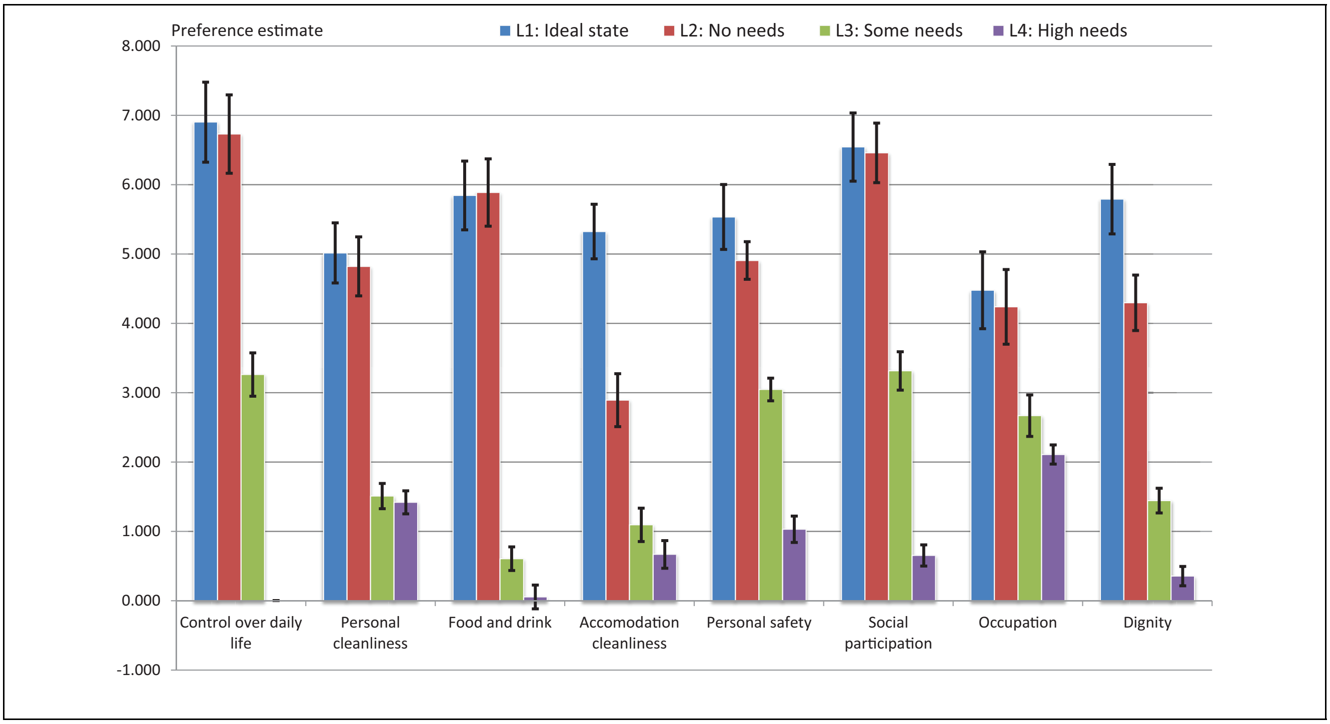


**Figure 1** Attribute-level coefficients and their 95% confidence intervals for the Finnish ASCOT for service users ($n = 32{,}000$).

only for the best or second-best choices. This finding confirmed that attributes in the profiles in a BWS design survey should be rotated to mitigate position bias that may influence the respondents' choice behavior and decisions, leading to invalid coefficient estimates.[64] In addition to the randomization of the attributes during the study design stage, researchers can control for the potential existence of position effects in the analysis data by including position-specific constants in the empirical model. Not accounting for the position effect can result in biased estimates for preference weights and may affect their validity if the estimates are used to provide policy recommendations.[64]

Two scale factors were influenced by education and health, which are recognized as being associated with cognitive functioning.[70,75] A short completion time of the BWS task as a significant scale factor could suggest that the respondents used a heuristic method to make choices quickly[76] or that they made a reduced effort while engaging in the choice tasks and considering the available alternatives properly. The learning effect has not been explored earlier in BWS studies, except in Nguyen et al.[24] In addition to two four-task sequences, we tested other sequential divisions of the choice tasks (e.g., the first two tasks vs. the later six tasks), but they were not statistically significant.

Similar to the position effects discussed above, the detected learning effect has more general implications for study designs and methods. When data are used to obtain utility estimates and to inform decision makers, researchers should pay attention to the ordering of the profiles in experiments in which choices are sequentially made.[32] In practice, to take into account the effects that respondent learning (or fatigue) might have on preference estimates, researchers can explicitly model learning (or fatigue) as a scale parameter, for example, by using the sequences of BWS tasks, as we have done in this study.

The scale factors were explicitly modelled to account for differences in random component variances between different groups of respondents or situations.[59] This also calls for approaches, by which researchers can separate scale heterogeneity from taste heterogeneity to derive accurate preference estimates (e.g., Flynn et al.[40]). In this respect, our modelling approach could be useful. We first studied taste heterogeneity (using the mixed logit) and then both taste and scale heterogeneity (using the G-MNL) assuming the values of coefficients are different to the worst and best choices. Having identified the scale factors, we used the taste-adjusted S-MNL model to derive the population-based preference estimates.

**Table 5** Final Preference Weights for the Finnish ASCOT for Service Users ($n = 32{,}000$)[a]

| Preference Weight | Level | Control Over Daily Life | Personal Cleanliness | Food and Drink | Accommodation Cleanliness | Personal Safety | Social Participation | Occupation | Dignity |
|---|---|---|---|---|---|---|---|---|---|
| *Panel 1.* Normalized values | 1. Ideal state | 1.000 | 0.727 | 0.853 | 0.649 | 0.771 | 0.802 | 0.948 | 0.839 |
| | 2. No needs | 0.975 | 0.699 | 0.847 | 0.614 | 0.419 | 0.711 | 0.936 | 0.622 |
| | 3. Some needs | 0.473 | 0.219 | 0.088 | 0.387 | 0.159 | 0.441 | 0.480 | 0.209 |
| | 4. High needs | 0.000 | 0.206 | 0.008 | 0.306 | 0.097 | 0.149 | 0.095 | 0.052 |
| *Panel 2.* Rescaled values | 1. Ideal state | 0.156 | 0.108 | 0.130 | 0.094 | 0.116 | 0.121 | 0.147 | 0.128 |
| | 2. No needs | 0.152 | 0.103 | 0.129 | 0.088 | 0.054 | 0.105 | 0.145 | 0.090 |
| | 3. Some needs | 0.063 | 0.018 | -0.005 | 0.048 | 0.008 | 0.058 | 0.065 | 0.017 |
| | 4. High needs | –0.020 | 0.016 | –0.019 | 0.034 | –0.003 | 0.006 | –0.003 | –0.011 |

[a]For the *food and drink* attribute, the current preference weight of level_1 was the originally estimated preference weight of level_2 and the current preference weight of level_2 was the originally estimated preference weight of level_1.

We have established a set of preference weights for the Finnish ASCOT for service users measure that are prerequisites for calculating social care QALYs (SC-QALYs) in Finland. Based on the relationship between BWS weights and time-trade off (TTO), Netten et al.[3] developed a formula for an English SC-QALY, with "0" equivalent to "being dead" and "1" being the "ideal" SCRQoL state. Possible SC-QALY scores in England range from –0.171 to + 1. Service users reported a significantly lower SCRQoL (score 0.73) than participants in the general population (score 0.86). Similar TTO study is out of scope of this article, but it may call for a new study in the future to derive a similar formula for a Finnish SC-QALY based on the association between Finnish TTO and BWS weights.

This study used preference data collected online. Evidence has indicated that modes of survey administration, such as internet-based surveys, might cause stronger fatigue effects and weaker learning effects,[31] but no notable comparable differences in the estimates from a model using online BWS data compared to those from a model using face-to-face interview data were observed.[77] Nonetheless, the learning effect was found in this study. The effect that the modes of survey administration may have on respondents' learning and fatigue presents an interesting area of future research in preference elicitation studies using the BWS method.

Our study has some limitations. First, since the respondents were recruited online, despite the quotas, the structure of the panel was not fully representative. This limitation also exists in other studies.[3,17] Nevertheless, we adjusted the preference weights as in the English and Austrian preference studies,[3,17] and computed the standard errors of the adjusted preference weights. Second, regardless of the exclusion of those with short completion times before the end of the data collection, the used survey administration method did not allow us to observe internal and external incentives or impetuses during the experiment, such as respondent behavior, burden and engagement, or changes in the task environment. However, we conducted face-to-face pretests to learn more about the participants' response behavior, which was also done in the Austrian study.[17]

To conclude, we have successfully established the preference weights for the Finnish ASCOT for service users instrument. Our contribution enlarges the number of valid measures that can be used to evaluate the capability-based QoL in a general population to consider the impact of social care interventions. The found learning effect calls for the development of study designs that reduce possible bias relating to preference uncertainty at the beginning of the test battery of BWS tasks. The finding also supports using a modelling strategy in which the sequence of tasks is explicitly modelled as a scale factor in an S-MNL model. Similarly, the attribute ordering effect calls for randomizing the items appearing in the choice list. The preference weights serve as a means to promote outcomes research in Finland and support Finnish policy makers in making evidence-based decisions regarding the use of resources for LTC services.

## Authors' Note

Ethical approval for this study was obtained from the Finnish Institute for Health and Welfare (THL), Finland, in January 2016.

## Acknowledgments

We would like to thank the participants at the NHESG Meeting 2017 (Helsinki), the Finnish Health Economics Day 2018 (Helsinki), and the EuHEA Conference 2018 (Maastricht) for the feedback.

## Data Availability Statement

The data used in this study are not publicly available. The Finnish Institute for Health and Welfare (THL) is committed to ensuring safe and secure use of the data.

## Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making Policy & Practice* website at https://journals.sagepub.com/home/mpp.

## ORCID iDs

Lien Nguyen https://orcid.org/0000-0002-8981-9041
Ismo Linnosmaa https://orcid.org/0000-0003-0178-8111

## References

1. Greve B. *Long-term Care for the Elderly in Europe. Development and Prospects*. Routledge; 2017.
2. Barbieri D, Ghibelli P. Social protection investment in long-term care. Formal versus informal long-term care: economic and social impacts [cited June 14, 2021]. Available from: http://sprint-project.eu/wp-content/uploads/2018/09/SPRINT_D4.4_Formal_vs_Informal-LTC_Econocmic_Social_Impacts.pdf
3. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16(16):1–166. doi:10.3310/hta16160
4. Dickens AP, Richards SH, Greaves CJ, Campbell JL. Interventions targeting social isolation in older people: a systematic review. *BMC Public Health*. 2011;11(1):647. doi:10.1186/1471-2458-11-647
5. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; 1996.
6. Bulamu NB, Kaambwa B, Ratcliffe J. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual Life Outcomes*. 2015;13(1):179. doi:10.1186/s12955-015-0372-8
7. Makai P, Brouwer WB, Koopmanschap MA, Stolk EA, Nieboer AP. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. *Soc Sci Med*. 2014;102:83–93. doi:10.1016/j.socscimed.2013.11.050
8. Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. *BMC Health Serv Res*. 2014;14(1):360. doi:10.1186/1472-6963-14-360
9. Luyten J, Kessels R, Goos P, Beutels P. Public preferences for prioritizing preventive and curative health care interventions: a discrete choice experiment. *Value Health*. 2015;18(2):224–33. doi:10.1016/j.jval.2014.12.007
10. Netten A, Ryan M, Smith P, et al. *The Development of a Measure of Social Care Outcome for Older People*. University of Kent; 2002.
11. Forder J, Netten A, Caiels J, Smith J, Malley J. *Measuring Outcomes in Social Care: Conceptual Development and Empirical Design*. University of Kent; 2007.
12. Netten A, Burge P, Malley J, et al. *Outcomes of Social Care for Adults (OSCA): Interim Findings*. University of Kent; 2009.
13. Sen A. Personal utilities and public judgements: or what's wrong with welfare economics. *Econ J*. 1979;89(355):537–58. doi:10.2307/2231867
14. Sen A. *Commodities and Capabilities*. North-Holland; 1985.
15. Kaambwa B, Gill L, McCaffrey N, et al. An empirical comparison of the OPQoL-Brief, EQ-5D-3 L and ASCOT in a community dwelling population of older people. *Health Qual Life Outcomes*. 2015;13(1):164. doi:10.1186/s12955-015-0357-7
16. Trukeschitz B. Worauf es letztlich ankommt: Ergebnisqualität in der Langzeitpflege und -betreuung. *Kurswechsel*. 2011;4:22–35.
17. Hajji A, Trukeschitz B, Malley J, et al. Population-based preference weights for the adult social outcomes toolkit (ASCOT) for service users for Austria: findings from a best-worst experiment. *Soc Sci Med*. 2020;250:112792. doi:10.1016/j.socscimed.2020.112792
18. Rostgaard T, Højmose M, Clement S, Rasmussen A. *Omsorgsbetinget livskvalitet og hjemmehjælp—en ASCOT-undersøgelse blandt hjemmehjælpsmodtagere*. Institut for Statskundskab, Aalborg Universitet; 2013.
19. Steffansson M, Pulliainen M, Kettunen A, Linnosmaa I, Halonen M. The association between freedom of choice and effectiveness of home care services. *Int J Integr Care*. 2016;16(1):5. doi:10.5334/ijic.2448
20. Cetrano G. Quality assessment of mental health and social care services. Findings from the European REFINEMENT project and an Italian multicentre study [thesis]. Verona, Italy: University of Verona; 2015.
21. Nakamura-Thomas H, Morikawa M, Moriyama Y, et al. Japanese translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT) in Japanese social service users. *Health Qual Life Outcomes*. 2019;17(1):59. doi:10.1186/s12955-019-1128-7
22. van Leeuwen KM, Malley J, Bosmans JE, et al. What can local authorities do to improve the social care-related quality of life of older adults living at home? Evidence from the Adult Social Care Survey. *Health Place*. 2014;29:104–13. doi:10.1016/j.healthplace.2014.06.004
23. Linnosmaa I, Nguyen L, Jokimäki H. *The EXCELC Project in Finland: The Main Descriptive Findings From Surveys Using the Adult Social Care Outcomes Toolkit (ASCOT)*. Finnish Institute for Health and Welfare (THL); 2020.
24. Nguyen L, Linnosmaa I, Jokimäki H, et al. Social care-related outcomes in Finland. Construct validity and

structural characteristics of the Finnish ASCOT measure with older home care users. *Health Soc Care Community*. 2021;29(3):712–28. doi:10.1111/hsc.13328
25. Diener E, Suh EM, Smith H, Shao L. National differences in reported subjective well-being: why do they occur? *Soc Indic Res*. 1995;34(1):7–32. doi:10.1007/BF01078966
26. Inglehart R, Klingemann HD. Genes, culture, democracy, and happiness. In: Diener E, Suh EM, eds. *Culture and Subjective Well-Being*. MIT Press; 2000. p 165–83.
27. Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna S. Cross-cultural adaptation of health measures. *Health Policy*. 1991;19(1):33–44. doi:10.1016/0168-8510(91)90072-6
28. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. *Qual Life Res*. 1993;2(6):467–75. doi:10.1007/bf00422221
29. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417–32. doi:10.1016/0895-4356(93)90142-n
30. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making*. 2001;21(1):7–16. doi:10.1177/0272989X0102100102
31. Savage SJ, Waldman DM. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *J Appl Econ*. 2008;23(3):351–71. doi:10.1002/jae.984
32. Carlsson F, Mørkbak MR, Olsen SB. The first time is the hardest: a test of ordering effects in choice experiments. *J Choice Model*. 2012;5(2):19–37. doi:10.1016/S1755-5345(13)70051-4
33. Day B, Bateman IJ, Carson RT, et al. Ordering effects and choice set awareness in repeat-response stated preference studies. *J Environ Econ Manage*. 2012;63(1):73–91. doi:10.1016/j.jeem.2011.09.001
34. Hess S, Hensher D, Daly AJ. Not bored yet—revisiting respondent fatigue in stated choice experiments. *Trans Res*. 2012;46:626–44. doi:10.1016/j.tra.2011.11.008
35. Finn A, Louviere JJ. Determining the appropriate response to evidence of public concern: the case of food safety. *J Public Policy Mark*. 1992;11(2):12–25. doi:10.1177/074391569201100202
36. Thurstone LL. A law of comparative judgement. *Psychol Rev*. 1994;101(2):266–70. doi:10.1037/0033-295X.101.2.266
37. Towers AM, Smith N, Palmer S, Welch E, Netten A. The acceptability and feasibility of using the Adult Social Care Outcomes Toolkit (ASCOT) to inform practice in care homes. *BMC Health Serv Res*. 2016;16(1):523. doi:10.1186/s12913-016-1763-1
38. Flynn TN, Marley AAJ. Best-worst scaling: theory and methods. In: Hess S, Daly A, eds. *Handbook of Choice Modelling*. Edward Elgar; 2014. p 1–342.
39. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: what it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171–89.
40. Flynn TN, Louviere JJ, Peters TJ, Coast J. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Soc Sci Med*. 2010;70(12):1957–65. doi:10.1016/j.jhealeco.2006.04.002
41. Potoglou D, Burge P, Flynn T, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med*. 2011;72(10):1717–27. doi:10.1016/j.socscimed.2011.03.027
42. Louviere J, Flynn T, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press; 2015. p 1–342.
43. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):259–67. doi:10.1586/erp.10.29
44. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICE-CAP—a capability instrument. Estimation of a UK general population tariff. *Health Econ*. 2015;24(3):258–69. doi:10.1002/hec.3014
45. Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. Nothing about us without us? A comparison of adolescent and adult health-state values for the child health utility-9D using profile case best–worst scaling. *Health Econ*. 2016;25(4):486–96. doi:10.1002/hec.3165
46. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3–13. doi:10.1016/j.jval.2012.08.2223
47. Bateman IJ, Carson RT, Day B, et al. *Economic Valuation With Stated Preference Techniques: A Manual*. Edward Elgar; 2002.
48. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4(1):8–11. doi:10.4103/0974-1208.82352
49. Statistics Finland. Population structure. Table: population according to age (1-year), sex, marital status and language by area 1990–2016 [cited December 11, 2017]. Available from: http://tilastokeskus.fi/til/vaerak/tau.html
50. Statistics Finland. Employment. Table: Population by main type of activity, sex, yearly age groups and year 1987–2016 [cited July 11, 2018]. Available from: http://www.stat.fi/til/tyokay/tau_en.html
51. Statistics Finland. Educational structure of population. Table: population aged 15 or over by level of education, municipality, gender and age 1970–2015 [cited July 11, 2018]. Available from: http://www.stat.fi/til/vkour/tau_en.html
52. Statistics Finland. Dwellings and housing conditions [cited October 11, 2017]. Available from: http://www.stat.fi/til/asas/tau_en.html
53. Statistics Finland. Population structure. Table: Population by religious community, age and sex in 2000 to 2016 [cited December 11, 2017]. Available from: http://tilastokeskus.fi/til/vaerak/tau.html

54. McFadden D. The measurement of urban travel demand. *J Public Econ*. 1974;3:303–28. doi:10.1016/0047-2727(74)90003-6
55. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. Academic Press; 1974. p 105–42.
56. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press; 2000.
57. Fiebig DG, Keane MP, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Market Sci*. 2010;29(3):393–421. http://www.jstor.org/stable/40608156
58. Lancsar E, Louviere J, Donaldson C, Currie G, Burgess L. Best worst discrete choice experiments in health: methods and an application. *Soc Sci Med*. 2013;76:74–82. doi:10.1016/j.socscimed.2012.10.007
59. Louviere J, Street D, Carson R, et al. Dissecting the random component of utility. *Mark Lett*. 2002;13(3):177-93. doi:10.1023/A:1020258402210
60. Keane MP. Modeling heterogeneity and state dependence in consumer choice behavior. *J Bus Econ Stat*. 1997;15(3):310–27. doi:10.1080/07350015.1997.10524709
61. Hess S, Train K. Correlation and scale in mixed logit models. *J Choice Model*. 2017;23:1–8. doi:10.1016/j.jocm.2017.03.001
62. Hall J, Fiebig DG, King MT, Hossain I, Louviere JJ. What influences participation in genetic carrier testing? Results from a discrete choice experiment. *J Health Econ*. 2006;25(3):520–37. doi:10.1016/j.jhealeco.2005.09.002
63. DeBell M, Krosnick JA. Computing weights for American National Election Study survey data. ANES Technical Report Series, No. nes012427 [cited January 11, 2021]. Available from: https://electionstudies.org/wp-content/uploads/2018/04/nes012427.pdf
64. Campbell D, Erdem S. Position bias in best-worst scaling surveys: a case study on trust in institutions. *Am J Agri Econ*. 2015;97(2):526–45. doi:10.1093/ajae/aau112
65. Kjær T, Bech M, Gyrd-Hansen D, Hart-Hansen K. Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Econ*. 2006;15(11):1217–28. doi:10.1002/hec.1117
66. Train KE. *Discrete Choice Methods With Simulation*. 2nd ed. Cambridge University Press; 2009.
67. ALOGIT. ALOGIT Software & Analysis Ltd. ALOGIT version 4.2 [cited September 14, 2017]. Available from: http://www.alogit.com/
68. Bierlaire M. BIOGEME: a free package for the estimation of discrete choice models. Paper presented at: Proceedings of the 3rd Swiss Transportation Research Conference; March 19–21, 2003; Ascona, Switzerland.
69. Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP capability index for older people. *Soc Sci Med*. 2008;67(5):874–82. doi:10.1016/j.socscimed.2008.05.015
70. Huynh E, Coast J, Rose J, Kinghorn P, Flynn T. Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. *Soc Sci Med*. 2017;189:114–28. doi:10.1016/j.socscimed.2017.07.012
71. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis: A Primer*. Cambridge University Press; 2005.
72. Jonker MF, Attema AE, Donkers B, Stolk EA, Versteegh MM. 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 Econ*. 2017;26(12):1534–47. doi:10.1002/hec.3445
73. Shiroiwa T, Moriyama Y, Nakamura-Thomas H, et al. Development of Japanese utility weights for the Adult Social Care Outcomes Toolkit (ASCOT) SCT4. *Qual Life Res*. 2020;29(1):253–63. doi:10.1007/s11136-019-02287-6
74. Swait J, Louviere J. The role of the scale parameter in the estimation and use of multinomial logit choice models. *J Marketing Res*. 1993;30(3):305–14. doi:10.1177/002224379303000303
75. Salthouse TA, Kausler DH, Saults JS. Age, self-assessed health status, and cognition. *J Gerontol*. 1990;45(4):P156–P160. doi:10.1093/geronj/45.4.p156
76. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):2. doi:10.1186/s13561-015-0079-x
77. Saloniki EC, Malley J, Burge P, et al. 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. *Qual Life Res*. 2019;28(8):2207–20. doi:10.1007/s11136-019-02172-2