ABSTRACT

Discrete choice experiments (DCEs) are frequently used in health economics to measure preferences for non-market goods. Best worst discrete choice experiment (BWDCE) has been proposed as a variant of the traditional "pick the best" approach. BWDCE, where participants choose the best and worst options, is argued to generate more precise preference estimates because of the additional information collected. However, the validity of the approach relies on two necessary conditions: (i) best and worst decisions provide similar information about preferences, and (ii) asking individuals to answer more than one choice question per task does not reduce data quality. Whether these conditions hold in empirical applications remains under researched. This is the first study to compare participants' choices across three experimental conditions: (i) BEST choices only, (ii) WORST choices only, and (iii) BEST & WORST choices (BWDCE). We find responses to worst choices are noisier. Implied preferences from the best only and worst only choices are qualitatively different, leading to different WTP values. Responses to BWDCE tasks have lower consistency and respondents are more likely to use simplifying decision heuristics. We urge caution in using BWDCE as an alternative to the traditional "pick the best" DCE.

1. Introduction

Discrete choice experiments (DCEs) are frequently used to elicit preferences for non-market goods (Clark et al, 2014; de Bekker-Grob et al, 2012; Hoyos, 2010). In a DCE survey, participants are asked to complete several choice tasks. Each choice task includes a limited number of choice options (e.g., Treatment A vs. Treatment B), which correspond to hypothetical multi-attribute descriptions of the product or service of interest (e.g., risk of side effects, effectiveness, and cost). The participants are then asked to choose their most preferred choice option. The observed choices allow estimation of the weights individuals attach to these different attributes. Estimated effects (or part-worth utilities) can then be used to derive economic measures such as marginal rate of substitution and more specifically, willingness-to-pay values (i.e., *how much individuals are willing to pay to improve the quality of the service by 1 unit?).*

Commented [A1]: **Reviewer's comment**: The authors should review the final manuscript for typos and grammar, as a few sentences have words missing, e.g. second sentence of introduction doesn't quite make sense as it stands. Should be "A DCE takes the form of a questionnaire…" and p10 last sentence "quality constraints" should be equality constraints" I think.

Authors' reply: We thank the reviewer for spotting these typo mistakes – The manuscript has now been carefully checked for typo mistakes.

DCEs make the identification of preferences possible by collecting a *large* number of observations regarding individuals' choice behaviour. This is typically achieved by asking a sample of participants to complete several choice tasks. However, asking participants too many choice tasks may reduce data quality (Bech et al, 2011), increasing the use of simplifying decision heuristics when answering choice questions (Cairns et al, 2002; Lagarde, 2013). These behavioural effects are of great concern for experimenters, threatening the validity of DCE results.

One attempt to mitigate these issues is to limit the number of choice tasks per participant, but to ask each participant to answer more choice questions per task. Best-worst discrete choice experiment (BWDCE), also known as BWS multi-profile or Best-Worse Scaling Case 3, was introduced into health economics to achieve this objective (Lancsar et al, 2013). BWDCE gathers extra preference information per choice set by asking respondents to choose the *best* option and the *worst* option[1]. In this regard, the BWDCE approach can be seen as an extension of the DCE approach. The popularity of this approach is growing. It has been applied to measure preferences in a range of areas (see **Table 1** for a review of BWDCE applications in the health literature).

The quantity and quality of information about individuals' preferences obtained from DCEs are corner-stones in making precise statistical inference and drawing valid and policy-relevant conclusions. Obtaining more information in DCE surveys reduces the

[1] Depending on the number of options presented in the choice task, additional best and worst choices may follow until an implied preference ordering over all alternatives within a choice task is achieved.

standard errors, narrows the confidence intervals around preference and welfare estimates, thus increasing the accuracy of parameter estimates. Given that the move from DCE to BWDCE is mainly motivated by the objective of measuring individuals' preferences more precisely, it is important to verify that the BWDCE approach can achieve this purpose. BWDCEs will provide more precise preferences estimates only if the best and worst choices generate the same information about individuals' preferences. In this case, BWDCE could be seen as a data augmentation procedure for DCE. Generating the same type of information about individuals' preferences implies that best and worst choices share the same determinants (i.e., same marginal sensitivity to changes in product attributes) and exhibit comparable levels of consistency (i.e., similar signal-to-noise ratio). These two conditions must hold to accept BWDCE as a valid extension of the DCE approach.

Whether these two conditions hold remains an open question. However, evidence from previous research raises concerns. Rather than directly asking participants to rank order the different choice options, the BWDCE approach asks respondents to identify the two extreme options, which should in principle correspond to the first and last ranked options. In this regard, the BWDCE approach can be seen as an implicit ranking exercise. Ben-Akiva et al (1991) investigated the reliability of stated preference ranking data and provided a clear demonstration of the potential for significant biases in simple pooling of ranking data. The stability of ranking information decreases with decreasing rank. Even after allowing for rank-specific scale and other bias parameters, the model combining all ranks was rejected. This finding undermined the validity of using ranking information to measure individuals' preferences and led researchers to focus on the standard "*pick the best only*" DCE approach.

Beyond concerns about ranking exercises, there are also good reasons to suspect that best and worst choices don't share the same determinants. In related research, there is evidence that selection and rejection decisions lead to different outcomes (Dhar & Wertenbroch, 2000; Laran & Wilcox, 2011; Shafir, 1993; Meloy & Russo, 2004). For example, in pairwise choice (i.e., option A vs. option B), there are differences in attributes weightings and amounts of attention paid to the different pieces of information between instructions to select the best option vs. reject the worst option.

In cases where best and worst choices differ in their determinants, a possible solution would be to exclude the worst choices from the analyses and treat the best choices as if they came from a standard DCE (Lancsar et al, 2013). This approach is valid only if asking participants to answer additional worst choice questions does not modify their response to best choice questions and/or undermine the quality of best choices (e.g. due to increased cognitive burden). To the best of our knowledge, this assumption has never been tested.

The objective of this paper is twofold. We first test the empirical validity of the BWDCE methodology by verifying whether the best and worst choices generate the same information about individuals' preferences. Second, we examine whether excluding worst choices from the estimation of preferences is a valid procedure to deal with incompatibilities of best and worst choices.

The remainder of this paper is divided into five sections. In Section 2 we describe the experimental design of the study and sampling strategy. Sections 3 and 4 test whether best and worst choices generate similar information about individuals' preferences. We first test whether the determinants of the decisions are the same (Section 3) and then investigate potential changes in consistency of decisions (Section 4). Section 5 examines the validity of excluding worst choices from the analyses and treating best choices as if they came from a standard DCE. We report methods and results for each Section separately. Section 6 discusses the implications of our results for the measurement of preferences and identifies avenues for future research.

2. Experimental design

2.1. Context

BWDCE was used to elicit preferences for supporting self-management of chronic pain. Detailed information on the study is available in Burton et al (2017). Attributes and levels are shown in **Table 2**. Chronic pain is commonly defined as any pain lasting more than 12 weeks. It may arise from an initial injury, such as a back sprain, or there may be an ongoing cause, such as illness. Chronic pain is often accompanied with other health problems and has a large impact on individuals' quality of life. Whilst chronic pain usually cannot be cured, it can be handled with self-management programmes (e.g., frequent workshops organised in community settings or medical facilities where trainers explain to participants how to reduce pain and improve functions, such that individuals can resume day-to-day activities).

2.2. The choice questionnaire

A statistically efficient design was employed to devise the choice tasks, minimising the sample size requirement for a given level of confidence (Rose & Bliemer, 2013). We replaced unknown parameters (i.e., *true* preferences and standard errors) with priors obtained from a pilot study (n=120). This resulted in 12 experimental choice tasks, each including three unlabelled choice options (**Figure 1**). We added two non-experimental tasks to familiarise the participants with the layout of the DCE (Task #1) and to test the monotonicity of choices (Task #14). The order of the choice tasks and choice options within the tasks was not randomised across the participants.

2.3. Experimental manipulation

The study is based on three experimental conditions:

- BOTH condition: Participants answer both the *best* and *worst* choice questions;
- BEST condition: Participants only answer *best* choice questions;
- WORST condition: Participants only answer *worst* choice questions.

In this article we refer to the two types of choices made by the participants as best and worst, but in line with the literature (see Table 1), these generic concepts were defined more precisely in the survey as "*like the most*" and "*like the least*" (Figure 1).

2.4. Recruitment and ethics

Participants were pseudo-randomly allocated across the three conditions. We first recruited a random sample of the general population for the BOTH condition. Four months later we recruited participants for the two remaining conditions by: (i) following the same recruitment method; (ii) recruiting a different sample from the target population; and (iii) randomly allocating participants to one of the two conditions.

We commissioned an online market research company (*ResearchNow!*) to recruit 500 respondents for the BOTH condition (517 achieved), 150 for BEST condition (156 achieved) and 150 for the WORST condition (155 achieved). The company targeted invitations to panel members whose profiles included any diagnosis associated with chronic pain. Invited panel members were screened for eligibility using the following criteria: (i) 16 years old or over; (ii) currently troubled by pain or discomfort, either all the time or on and off; and (iii) had pain or discomfort for more than three months.

A copy of the questionnaire for the BOTH condition, scripted by *ResearchNow!* is provided in the supplementary information. The questionnaires for the BEST and WORST conditions were identical in all ways other than the choice questions.

Characteristics of participants are shown in **Table 3**. The three samples are similar in terms of relationship, educational level, employment, and health. The BEST and WORST conditions differ in terms of age. The BEST and BOTH conditions differ in terms of gender, income level and age.

The study was approved by the North of Scotland Research Ethics Service (Reference 14/NS/0075). Participants in the developmental stages all provided informed consent to take part. Consent for participants was managed by the market research company.

3. Do best and worst choices share the same determinants?

3.1. Methods

We first compare respondents' preferences across the BEST (n=156; #obs=1,872) and WORST (n=155; #obs=1,860) experimental conditions. Given preference estimates are confounded with the scale parameter, which is inversely related to the error variance, we cannot directly compare parameter estimates. We thus compare willingness-to-pay (WTP) values (comparing ratios eliminates this problem). We specified a WTP-space multinomial logit (MNL) model (Scarpa et al, 2008), estimating WTP values for the four qualitative attributes (i.e., information, situation, living well, communication). We included interaction effects between the WTP parameters and type of experimental condition (i.e., BEST vs. WORST) to determine whether valuation of attributes significantly differ across the two types of choices. We thus specify the following model:

$$U_{ntj} = \lambda_{ntj} V_{ntj} + \varepsilon_{ntj} \qquad \text{(Eq. 1)}$$

$$\varepsilon_{ntj} \sim \text{iid EV1} \qquad \text{(Eq. 2)}$$

$$\lambda_{ntj} = \frac{\pi}{\sigma_\varepsilon \sqrt{6}} \qquad \text{(Eq. 3)}$$

$$V_{ntj} = \beta_1 \mathrm{OPT2}_{ntj} + \beta_2 \mathrm{OPT3}_{ntj} - \mathrm{COST}_{ntj}(\beta_3 + \beta_{4:7}\mathrm{AGE}_n) + \beta_3 \Big[\gamma_1 \mathrm{INFO}_{ntj} + \gamma_2 \mathrm{SITU}_{ntj} + \gamma_3 \mathrm{LIVE}_{\mathrm{WELL}_{ntj}} + \gamma_4 \mathrm{COMM}_{ntj} + \mathrm{BEST}_{ntj}\Big(\delta_1 \mathrm{INFO}_{ntj} + \delta_2 \mathrm{SITU}_{ntj} + \delta_3 \mathrm{LIVE}_{\mathrm{WELL}_{ntj}} + \delta_4 \mathrm{COMM}_{ntj}\Big)\Big] \qquad \text{(Eq. 4)}$$

$$P_{ntj} = \frac{\exp(V_{ntj})}{\sum_j \exp(V_{ntj})} \qquad \text{(Eq. 5)}$$

Where the utility (U) of the choice option (j) faced by respondent (n) in choice task (t) depends on a systematic component (V) which can be explained and a stochastic component (ε) which is unobservable. Following the choices modelling literature, this stochastic component is typically assumed to be independently and identically distributed type I extreme value, leading thus to the so-called multinomial logit (MNL) model. Because of this stochastic component, one can only predict the probability of an option to be chosen (P) that is to yield the highest level of utility. (λ_{ntj}) is a scale parameter which is inversely related to the variance of the stochastic component (σ_ε). As this scale parameter is perfectly confounded with preference parameters, it is typically assumed to be equal to 1 for identification purpose (constraining thus σ_ε to become a fixed quantity) (Train, 2009).

In this model, the parameters of interest are ($\gamma_{1:4}$) which correspond to the WTP estimates in the worst condition, ($\delta_{1:4}$) which measure the marginal change in WTP values when moving from WORST to BEST conditions. If the best and worst choices share exactly the same determinants, all four interaction effects should be null (i.e., H_0: $\delta_{1:4} = 0$). If best and worst result in different preferences then H_1: $\delta_{1:4} \neq 0$. We control for participants' age (given it was significantly different between the BEST and WORST conditions).

3.2. Results

Results are presented in **Table 4**. Allowing for interaction effects between type of choice and WTP improved model fit (Log-likelihood ratio test: Deviance = 33; DF = 4; P < 0.001). Two interaction effects reach significance, indicating WTP differ between best and worst choices. For example, the Information attribute has a WTP value twice as large in the BEST vs. WORST condition (£16.1 vs. £8.5)[2].

4. Are best and worst choices equally consistent?

4.1. Methods

We again compare respondents' choices between the BEST (n=156; #obs=1,872) and WORST (n=155; #obs=1,860) experimental conditions. As noted above, parameter estimates confound *true* preferences and errors variance. Changes in error variance

[2] As Information is interacted with type of choice (BEST), its main effect becomes the WTP for the reference category (i.e., BEST = 0 => WORST choices) and the interaction effect captures the marginal effect on the WTP of moving from WORST to BEST. Therefore, the WTP for WORST is 8.582. The marginal effect is 7.524, such that WTP for BEST is 8.582 + 7.524 = 16.106.

represent differences in choice consistency (DeShazo and Fermo 2002; Börger 2016). If respondents make more random decisions, errors variance increases. To test this across our best and worst conditions we specify a heteroskedastic MNL model, allowing the error variance to depend on the type of choices (Hole, 2006; Swait & Adamowicz, 2001)[3].

$$V_{ntj} = \beta_1 \mathrm{OPT2}_{ntj} + \beta_2 \mathrm{OPT3}_{ntj} + \beta_3 \mathrm{INFO}_{ntj} + \beta_4 \mathrm{SITU}_{ntj} + \beta_5 LIVE_{ntj} + \beta_6 \mathrm{COMM}_{ntj} + \beta_7 \mathrm{COST}_{ntj} \quad \text{(Eq. 6)}$$

$$\lambda_{ntj} = \exp\left(\alpha_1 \mathrm{BEST}_{ntj} + \alpha_{2:5} \mathrm{AGE}_{ntj}\right) \quad \text{(Eq. 7)}$$

We again control for age differences across experimental conditions. The ($\beta_{1:2}$) parameters are constants. The ($\beta_{3:7}$) are preference parameters capturing the marginal sensitivity to changes in the five attributes: Information (INFO), situation (SITU), living well (LIVE), communication (COMM), and cost (COST). No differences between BEST and WORST conditions imply same level of choice consistency (i.e., H_0: $\alpha_1 = 0$). However Lancsar et al (2013) obtained larger preference estimates when analysing best choices only compared to jointly analysing best and worst choices, suggesting larger error variance for worst choices. We thus expect worst choices to be less consistent than best choices (i.e., H_1: $\alpha_1 > 0$).

4.2. Results

Results are presented in **Table 5**. Preferences for the five attributes are in line with *a priori* assumptions (i.e., the four personalisation attributes have a positive impact on utility, and utility is decreasing with increased cost). Best choices are more consistent than worst choices, as indicated by the positively significant scale parameter ($\alpha_1 = 0.183$, $p < 0.001$)). Thus, we reject the assumption that best and worst choices are equally consistent.

5. Do worst choices negatively influence the quality of best choices?

5.1. Methods

We examine the influence of asking participants to make worst choices (in addition to the best) on the quality of best choices. For this purpose, we compare best choices between the BEST (n=156; #obs=1,872) and BOTH (n=517; #obs=6,204) experimental

[3] We also estimated a MNL model accounting for the panel nature of the data by adding an individual-level error terms: $u_n \sim \mathrm{Normal}(0; \sigma)$. However, the ($\sigma$) parameter was not significant and the model failed to outperform the initial MNL model. Results can be obtained from the corresponding author upon request.

conditions. We first investigate consistency of the choices i.e., *does being asked to answer an additional worst choice question negatively impact the consistency of the best choices?* We then explore potential changes in the underlying decision rules i.e., *do participants approach the best choice question the same way when being also asked to answer an extra worst choice question?* In both analyses we control for age, gender, and income differences across experimental conditions.

<u>Influence on choices consistency</u>

We investigate the effect of asking participants to answer a worse choice question per task on the consistency of best choices by estimating a heteroscedastic MNL model allowing the errors variance to differ between the BOTH and BEST conditions.

$$\lambda_{ntj} = \exp\left(\alpha_1 \mathrm{BOTH}_{ntj} + \alpha_2 \mathrm{FEMALE}_{ntj} + \alpha_{3:5} \mathrm{INCOME}_{ntj(1:3)} + \alpha_{6:9} \mathrm{AGE}_{ntj(1:4)}\right) \quad \text{(Eq. 8)}$$

Where (α_1) captures the effect on scale of being in the BOTH condition relative to BEST condition. We expect to find a negative effect ($\alpha_1 < 0$), meaning that best choices are less consistent in the BOTH condition (as a consequence of increased cognitive burden).

<u>Influence on decision rules</u>

Participants may respond to changes in the cognitive burden of the choice tasks (due to answering an extra choice question) by adjusting the amount of information they consider when making their choices. We approximate information processing strategies with attribute non-attendance (ANA). This describes a form of information processing in which a piece of information (or attribute) is either considered or ignored. In the DCE literature, ANA has been investigated either by directly asking individuals to *state* which attributes they have ignored or considered, or *inferred* by using ANA choice models (Campbell et al, 2011; Scarpa et al, 2013; Hole et al, 2013). We followed this latter approach. Given that the number of possible ANA patterns grows quickly with the number of attributes[4], we limit our analysis to four main ANA patterns or information processing strategies, thus improving model tractability.

The first benchmark pattern corresponds to a case where all the attributes were attended or considered (i.e., full information processing, FIP). The second pattern

[4] Our choice experiment includes five attributes allowing thus for 32 different ANA patterns.

defines the opposite case where none of the attributes were attended (i.e., null information processing, NIP) and then individuals' choices would be made randomly. The two remaining patterns represent intermediate cases (i.e., partial information processing, PIP) where the individuals only considered either the cost attribute (PIP_{COST}) or the four quality-related attributes ($PIP_{QUALITY}$). We accommodate these four ANA patterns in a constrained latent class logit (LCL) model. The objective of this LCL model is to determine whether having to answer two choice questions instead of one makes participants more likely to use strategies other than FIP. As the LCL model also allows to model class membership (i.e., the probability of adopting any of the four information processing strategies), we determine whether respondents from the BOTH condition are more likely to belong to one of the non-FIP classes.

$$\mathrm{FIP} = \beta_{11}\mathrm{OPT2} + \beta_{12}\mathrm{OPT3} + \beta_{13}\mathrm{INFO} + \beta_{14}\mathrm{SITU} + \beta_{15}\mathrm{LIVE} + \beta_{16}\mathrm{COMM} + \beta_{17}\mathrm{COST} \quad \text{(Eq. 9)}$$

$$\mathrm{PIP}_{\mathrm{quality}} = \beta_{21}\mathrm{OPT2} + \beta_{22}\mathrm{OPT3} + \beta_{23}\mathrm{INFO} + \beta_{24}\mathrm{SITU} + \beta_{25}\mathrm{LIVE} + \beta_{26}\mathrm{COMM} \quad \text{(Eq. 10)}$$

$$\mathrm{PIP}_{\mathrm{cost}} = \beta_{31}\mathrm{OPT2} + \beta_{32}\mathrm{OPT3} + \beta_{37}\mathrm{COST} \quad \text{(Eq. 11)}$$

$$\mathrm{NIP} = \beta_{41}\mathrm{OPT2} + \beta_{42}\mathrm{OPT3} \quad \text{(Eq. 12)}$$

These different specifications of the utility functions constrain some parameters to be null. By doing so, we assume that some respondents have not considered these attributes when making their choices. For example, in the $PIP_{quality}$ function (Eq. 10), the cost attribute is supposed to be ignored, and then the corresponding parameter is constrained to be null ($\beta_{27} = 0$). In addition to these nullity constraints, we also constrain the remaining preference parameters to be the same across the four classes (i.e., $\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41}$; $\beta_{12} = \beta_{22} = \beta_{32} = \beta_{42}$; $\beta_{13} = \beta_{23}$; $\beta_{14} = \beta_{24}$; $\beta_{15} = \beta_{25}$; $\beta_{16} = \beta_{26}$; $\beta_{17} = \beta_{37}$). These equality constraints have been added to reduce the confounding effect of preferences heterogeneity. Without these equality constraints, the LCL model would capture differences in decision rules and in preferences.

$$\mathrm{Membership} = \alpha_{1:3} + \alpha_{4:6}\mathrm{BOTH}_{\mathrm{n}} + \alpha_{7:9}\mathrm{FEMALE}_{\mathrm{n}} + \alpha_{10:18}\mathrm{INCOME}_{\mathrm{n}(1:3)} + \alpha_{19:30}\mathrm{AGE}_{\mathrm{n}(1:4)} \quad \text{(Eq. 11)}$$

In the membership function, the ($\alpha_{4:6}$) parameters capture the effect of the experimental condition. We expect participants to be more likely to use a non-FIP strategy in the BOTH condition ($\alpha_4 > 0$; $\alpha_5 > 0$; $\alpha_6 > 0$).

5.2. Results

Influence on choices consistency

Results for the heteroscedastic MNL model are presented in **Table 6**. This model significantly outperforms its homoscedastic counterpart assuming similar errors variance for the best choices in the BEST and BOTH conditions (LR test: Deviance = 83.6; DF = 9; P < 0.001). As expected the (α_1) parameter is negative and significant, indicating that adding a worst choice question makes answers to the best choice questions less consistent.

Influence on decision rules

The results of the LCL model are presented in **Table 7**. Allowing for different information processing strategies improves model performance (MNL_{LL} = -7,530.6 vs. LCL_{LL} = -7,110.4) even after adjusting for the number of model parameters (MNL_{BIC} = 15,124.1 vs. LCL_{BIC} = 14,553.8). Regarding class membership (α) parameters, participants in the BOTH condition are significantly more likely to adopt a $PIP_{QUALITY}$ and NIP strategy. By using the ($\hat{\alpha}$) estimates, it is possible to compute each respondent's probability of belonging to one of the four information processing classes[5]. This analysis shows that the BOTH condition was associated with a significantly higher share for the $PIP_{QUALITY}$ class compared to the BEST (i.e., 64.1% → 91.7%). This increase comes at the expense of a large decrease in the FIP class share (i.e., 35.9% → 1%). The shares for the NIP and PIP_{COST} classes remain low and comparable between the BOTH and BEST conditions (i.e., NIP: 0% → 4.1%; PIP_{COST}: 0% → 3.3%). These results indicate that participants to a BWDCE (BOTH condition) are more likely to adopt simplifying decision rules than participants to a standard DCE (BEST condition).

6. Discussion

In the DCE literature, the BWDCE (also known as BWS multi-profile or BWS case 3) has been used as an extended version of the standard "pick the best only" DCE approach. It is argued that BWDCE measures individuals' preferences more precisely. However, one necessary condition to achieve this is that best and worst choices generate the same information about individuals' preferences. We test for the first time the empirical validity of this assumption using an appropriate (split-sample)

[5] We used the following decision rule for class allocation: a respondent was allocated to the class for which s/he has the highest probability of belonging.

experimental design. We show that best and worst choices do not generate the same information about preferences. These two types of decisions significantly differ in their determinants (i.e., importance given to the different product attributes) and in their level of consistency, with worst choices being noisier than best choices. These results question the standard practice of pooling best and worst data in discrete choice models.

The differences in determinants suggest that the choice model should incorporate two different utility functions (i.e., one to explain the best choices and another to explain the worst choices). However, this would also lead to two different sets of preference estimates and it is unclear how useful this type of result would be from a policy-making perspective (i.e., which set of preference should be used to make predictions about individuals future health decisions? What if the two sets lead to different recommendations/conclusions?). In a previous BWDCE, Lancsar et al (2013) also found larger preference estimates when using best choices alone compared to a model combining both best and worst choices, suggesting a larger error variance for the worst choices. However, the best and worst choices came from the same questionnaire thus limiting the scope of their results. Xie et al (2014) showed that standard "pick the best only" DCE and BWDCE perform equally well, but found that DCE choices were easier and shorter to complete. The authors concluded that the DCE was more feasible and reliable than the BWDCE in valuing EQ-5D-5L health states. As in previous BWDCE applications, a limitation of their study is that the empirical analyses relied on the assumption that best and worst choices generate the same information about individuals' preferences and could thus be pooled. Our study provides empirical evidence suggesting that this strategy is not valid.

One practical solution to this problem of choices incompatibilities would be to exclude the worst choices from the analyses, thus treating best choices as if they come from a standard DCE (in which participants only choose their preferred option). We tested the validity of this assumption, by comparing the quality of best choices in a BWDCE versus DCE survey, and found it was not verified empirically. The consistency of the best choices decreased and participants were more likely to adopt simplifying decision rules in the BWDCE survey, questioning thus the external validity of the estimated preferences.

Some questions are left unanswered in our study. We found a lower quality level for the worst choices compared to the best. However, this result might not be generalizable to other settings. In some cases, such as research on health states and quality of life, making worst choices may be easier and more relevant (Burr et al, 2007; Ryan et al, 2006) and therefore generate better quality data. When designing a DCE, the appropriateness of asking best and/or worst choice questions should be explored at the piloting stage using qualitative approaches (Ryan et al, 2009).

We did not consider the issue of heterogeneity in the views regarding best and worst choices. The BWDCE approach could be more appropriate for participants who are better at determining what they don't like (worst choice) rather than what they like (best choice). Future work could explore this by collecting information on the personality type of respondents. For example, individuals who tend "*to see the glass as half empty rather than half full*" may be better placed to answer worst choice questions.

There might be an interaction between the design properties and the quality of the best and worst choices. We used a statistically efficient design to increase the amount of information about preferences obtained from each choice. This type of design is typically associated with a higher level of cognitive difficulty for the participants because it increases the similarity between the choice options, thus leading to more complex trade-offs (Yao et al, 2015; Reed Johnson et al, 2013). The BWDCE method can be seen as a variant of the ranking approach, taking advantage of human ability to better identify extreme events (e.g., highly desirable vs highly undesirable options). Therefore an experimental design maximising the statistically efficiency of the DCE by making the choice options more similar might be less compatible with a BWDCE approach.

Finally, a variant of the BWDCE has been proposed, asking individuals to first choose their most preferred option (1st best) and then their next preferred option (2nd best) (Lancsar et al, 2017; Ghijben et al, 2014). However it is not clear how this "best-best" approach would differ from traditional ranking tasks, with their associated limitations (Ben-Akiva et al, 1991). This sequential approach may help to break down a complex decision problem (i.e., to rank all choice options in terms of desirability) into more manageable tasks, thus yielding better quality data. This remains an empirical question.

Consideration should be given to the relevance of our findings to other types of BWS experiments (i.e., cases 1 and 2). In the more commonly use BWS case 2 approach, participants face one profile at a time and are asked to choose its best and worst features (i.e., most and least desirable attributes' levels). The relevance of our findings for BWS case 2 studies depend on the modelling strategy adopted. Rather than directly analysing the probability of being selected as best (worst) attributes' level, studies have typically used the Maximum Difference (MaxDiff) model to analyse BW responses. This approach models the probability of picking a "best-worst" pair of attributes' levels among all possible pairs. Thus, the MaxDiff approach is less concerned with differences in determinants of best and worst choices. However, the MaxDiff approach does not match the *true* data generating process (i.e., it is unlikely to describe how respondents have completed the choice tasks) and therefore one might want to consider a direct analysis of the best and worst choices. In this case, differences in determinants of best and worst choices would also be a central issue for the analysis of BWS case 2 data. This comment also applies to BWS case 1 studies.

Commented [A2]: __Reviewer's comment__: I am not entirely sure I agree with everything in the new paragraph at line 10, p14. I think the question of whether there is a different best and worst choice process is still relevant in the BWS Case 2 context. The models are still based on the assumption the process is the same.

__Authors' reply__: We have addressed this comment by revising the paragraph – We agree with the editor that differences of determinants between best and worst choices is an issue for all three variants of the BWS approach (case 1, case 2 and case 3) – However this issue can be artificially attenuated by using different modelling strategies such as MaxDiff (as now explained in the text).

Our study is not free from limitations. We recruited participants at two points in time. Whilst we adopted the same recruitment method, a short time elapsed between experiments, and our analysis allowed for differences in observable characteristics, we cannot rule out sampling effects. Also it is possible but unlikely that some participants have answered two different versions of the questionnaire. In a different project, we found that 17.6% of the participants, who are also members of an online panel, already took part in a DCE survey before (i.e., "*In this survey we asked you which dental care packages you preferred. Have you ever completed a similar survey (where you were asked to make choices between alternative goods or services) in a health context*"). Assuming some participants took part in the two different versions of our questionnaire (i.e., BEST/WORST and BOTH), the effect on the study results should be limited. A potential learning or experience effect would work against our main research conjecture (i.e., best and worst choices differ in their determinants) because respondents would try to be consistent in their decisions, attenuating thus the effect of the experimental manipulation.

Second, in our analysis of information processing rules, we specified a latent class logit (LCL) model allowing for different types of decision rules. However these decision rules are likely to be confounded with differences in preferences. That is, it is practically impossible to differentiate between a null weight for the cost attribute due to cost being ignored vs. participants having very low cost sensitivity (Hess et al, 2013; Alemu et al, 2013).

Third, when modelling the best and worst choices we assumed simultaneity in the decision-making, such that both types of choices should have been made at the same time rather than sequentially. We also tested choice models allowing for a sequential decision-making (see **online supporting information**), this did not improve model fit.

6. Conclusion

Our results challenge the current view that BWDCE can be used as an alternative to standard DCEs to measure individuals' preferences more precisely. The extra information obtained from the worst choices was found to be different from that obtained from the best choices. More specifically, best and worse choices generated different WTP values for individual attributes and best choices were more consistent than worst choice. Best choices observed in a BWDCE appeared to be less consistent and individuals were more likely to adopt simplifying decision heuristics. We urge caution in using BWDCE as an alternative to the traditional "pick the best" DCE.

References

Alemu, Mohammed Hussen, Morten Raun M?rkbak, S?ren B?ye Olsen, and Carsten Lynge Jensen. 2013. “Attending to the Reasons for Attribute Non-Attendance in Choice Experiments.” *Environmental and Resource Economics* 54 (3): 333–59. https://doi.org/10.1007/s10640-012-9597-8.

Bech, Mickael, Trine Kjaer, and Jørgen Lauridsen. 2011. “Does the Number of Choice Sets Matter? Results from a Web Survey Applying a Discrete Choice Experiment.” *Health Economics* 20 (3): 273–86. https://doi.org/10.1002/hec.1587.

Bekker-Grob, Esther W. de, Mandy Ryan, and Karen Gerard. 2012. “Discrete Choice Experiments in Health Economics: A Review of the Literature.” *Health Economics* 21 (2): 145–72. https://doi.org/10.1002/hec.1697.

Ben-Akiva, Moshe, Takayuki Morikawa, and Fumiaki Shiroishi. 1991. “Analysis of the Reliability of Preference Ranking Data.” *Journal of Business Research* 23 (3): 253–68. https://doi.org/10.1016/0148-2963(91)90033-T.

Börger, Tobias. 2016. “Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment.” *Environmental and Resource Economics* 65 (2): 389–413. https://doi.org/10.1007/s10640-015-9905-1.

Burr, Jennifer M., Mary Kilonzo, Luke Vale, and Mandy Ryan. 2007. “Developing a Preference-Based Glaucoma Utility Index Using a Discrete Choice Experiment:” *Optometry and Vision Science* 84 (8): 797–808. https://doi.org/10.1097/OPX.0b013e3181339f30.

Burton, Christopher D, Vikki A Entwistle, Alison M Elliott, Nicolas Krucien, Terry Porteous, and Mandy Ryan. 2017. “The Value of Different Aspects of Person-Centred Care: A Series of Discrete Choice Experiments in People with Long-Term Conditions.” *BMJ Open* 7 (4): e015689. https://doi.org/10.1136/bmjopen-2016-015689.

Cairns, John, Marjon van der Pol, and Andrew Lloyd. 2002. “Decision Making Heuristics and the Elicitation of Preferences: Being Fast and Frugal about the Future.” *Health Economics* 11 (7): 655–58. https://doi.org/10.1002/hec.720.

Campbell, Danny, David A. Hensher, and Riccardo Scarpa. 2011. “Non-Attendance to Attributes in Environmental Choice Analysis: A Latent Class Specification.” *Journal of Environmental Planning and Management* 54 (8): 1061–76. https://doi.org/10.1080/09640568.2010.549367.

Clark, Michael D., Domino Determann, Stavros Petrou, Domenico Moro, and Esther W. de Bekker-Grob. 2014. “Discrete Choice Experiments in Health Economics: A Review of the Literature.” *PharmacoEconomics* 32 (9): 883–902. https://doi.org/10.1007/s40273-014-0170-x.

DeShazo, J.R., and German Fermo. 2002. “Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency.” *Journal of Environmental Economics and Management* 44 (1): 123–43. https://doi.org/10.1006/jeem.2001.1199.

Dhar, Ravi, and Klaus Wertenbroch. 2000. “Consumer Choice Between Hedonic and Utilitarian Goods.” *Journal of Marketing Research* 37 (1): 60–71. https://doi.org/10.1509/jmkr.37.1.60.18718.

Ghijben, Peter, Emily Lancsar, and Silva Zavarsek. 2014. “Preferences for Oral Anticoagulants in Atrial Fibrillation: A Best–Best Discrete Choice Experiment.” *PharmacoEconomics* 32 (11): 1115–27. https://doi.org/10.1007/s40273-014-0188-0.

Hess, Stephane, Amanda Stathopoulos, Danny Campbell, Vikki O'Neill, and Sebastian Caussade. 2013. "It's Not That I Don't Care, I Just Don't Care Very Much: Confounding between Attribute Non-Attendance and Taste Heterogeneity." *Transportation* 40 (3): 583–607. https://doi.org/10.1007/s11116-012-9438-1.

Hole, Arne Risa. 2006. "Small-Sample Properties of Tests for Heteroscedasticity in the Conditional Logit Model." *Economics Bulletin* 3: 1–14.

Hole, Arne Risa, Julie Riise Kolstad, and Dorte Gyrd-Hansen. 2013. "Inferred vs. Stated Attribute Non-Attendance in Choice Experiments: A Study of Doctors? Prescription Behaviour." *Journal of Economic Behavior & Organization* 96 (December): 21–31. https://doi.org/10.1016/j.jebo.2013.09.009.

Hoyos, David. 2010. "The State of the Art of Environmental Valuation with Discrete Choice Experiments." *Ecological Economics* 69 (8): 1595–1603. https://doi.org/10.1016/j.ecolecon.2010.04.011.

Lagarde, Mylene. 2013. "INVESTIGATING ATTRIBUTE NON-ATTENDANCE AND ITS CONSEQUENCES IN CHOICE EXPERIMENTS WITH LATENT CLASS MODELS: ATTRIBUTE NON-ATTENDANCE IN CHOICE EXPERIMENTS." *Health Economics* 22 (5): 554–67. https://doi.org/10.1002/hec.2824.

Lancsar, Emily, Denzil G. Fiebig, and Arne Risa Hole. 2017. "Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software." *PharmacoEconomics* 35 (7): 697–716. https://doi.org/10.1007/s40273-017-0506-4.

Lancsar, Emily, Jordan Louviere, Cam Donaldson, Gillian Currie, and Leonie Burgess. 2013. "Best Worst Discrete Choice Experiments in Health: Methods and an Application." *Social Science & Medicine* 76 (January): 74–82. https://doi.org/10.1016/j.socscimed.2012.10.007.

Laran, Juliano, and Keith Wilcox. 2011. "Choice, Rejection, and Elaboration on Preference-Inconsistent Alternatives." *Journal of Consumer Research* 38 (2): 229–41. https://doi.org/10.1086/659040.

Meloy, Margaret G, and J.Edward Russo. 2004. "Binary Choice under Instructions to Select versus Reject." *Organizational Behavior and Human Decision Processes* 93 (2): 114–28. https://doi.org/10.1016/j.obhdp.2003.12.002.

Reed Johnson, F., Emily Lancsar, Deborah Marshall, Vikram Kilambi, Axel Mühlbacher, Dean A. Regier, Brian W. Bresnahan, Barbara Kanninen, and John F.P. Bridges. 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. https://doi.org/10.1016/j.jval.2012.08.2223.

Rose, John M., and Michiel C. J. Bliemer. 2013. "Sample Size Requirements for Stated Choice Experiments." *Transportation* 40 (5): 1021–41. https://doi.org/10.1007/s11116-013-9451-z.

Ryan, Mandy, Ann Netten, Diane Sk?tun, and Paul Smith. 2006. "Using Discrete Choice Experiments to Estimate a Preference-Based Measure of Outcome?An Application to Social Care for Older People." *Journal of Health Economics* 25 (5): 927–44. https://doi.org/10.1016/j.jhealeco.2006.01.001.

Ryan, Mandy, Verity Watson, and Vikki Entwistle. 2009. "Rationalising the Irrational: A Think Aloud Study of Discrete Choice Experiment Responses." *Health Economics* 18 (3): 321–36. https://doi.org/10.1002/hec.1369.

Scarpa, R., R. Zanoli, V. Bruschi, and S. Naspetti. 2013. “Inferred and Stated Attribute Non-Attendance in Food Choice Experiments.” *American Journal of Agricultural Economics* 95 (1): 165–80. https://doi.org/10.1093/ajae/aas073.

Scarpa, Riccardo, Mara Thiene, and Kenneth Train. 2008. “Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps.” *American Journal of Agricultural Economics* 90 (4): 994–1010. https://doi.org/10.1111/j.1467-8276.2008.01155.x.

Shafir, Eldar. 1993. “Choosing versus Rejecting: Why Some Options Are Both Better and Worse than Others.” *Memory & Cognition* 21 (4): 546–56. https://doi.org/10.3758/BF03197186.

Swait, Joffre, and Wiktor Adamowicz. 2001. “Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice.” *Organizational Behavior and Human Decision Processes* 86 (2): 141–67. https://doi.org/10.1006/obhd.2000.2941.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge ; New York: Cambridge University Press.

Whitty, Jennifer A., and Ana Sofia Oliveira Gonçalves. 2018. “A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best–Worst Scaling for Eliciting Preferences in Healthcare.” *The Patient - Patient-Centered Outcomes Research* 11 (3): 301–17. https://doi.org/10.1007/s40271-017-0288-y.

Xie, Feng, Eleanor Pullenayegum, Kathryn Gaebel, Mark Oppe, and Paul F. M. Krabbe. 2014. “Eliciting Preferences to the EQ-5D-5L Health States: Discrete Choice Experiment or Multiprofile Case of Best?Worst Scaling?” *The European Journal of Health Economics* 15 (3): 281–88. https://doi.org/10.1007/s10198-013-0474-3.

Yao, Richard T., Riccardo Scarpa, John M. Rose, and James A. Turner. 2015. “Experimental Design Criteria and Their Behavioural Efficiency: An Evaluation in the Field.” *Environmental and Resource Economics* 62 (3): 433–55. https://doi.org/10.1007/s10640-014-9823-7.

Table 1 Studies using the best-worst discrete choice experiments (BWDCEs) in health

Author (Year)	Topic	Sample size	# tasks	# options	Task format	Wording of the choice questions
Brown (2011)	Haemophilia treatment	53	12	3	First best; First worst	Q1. Which treatment are you most likely to use? Q2. Which treatment are you least likely to use?
Cameron (2013)	HIV vaccine	324	1	8	First best; First worst; Ranking of remaining 6 options	*Not detailed*
Gallego (2015)	Job characteristics	165	7	3	First best; First worst	Q1. Of these jobs, which one would most likely keep you practising in a rural area? Q2. Of these jobs, which one would least likely keep you practising in a rural area?
Hoek (2011)	Packaging for tobacco control	292	13	4	First best; First worst	Q1. Which pack would you be most likely to choose? Q2. Which pack would you be least likely to choose?
Lancsar (2013)	Treatment of cardiac arrest	898	16	5	First best; First worst; Second best; Second worst	Q1. Which option do you prefer most? Q2. Which option do you prefer least? Q3. Which of the three remaining options do you prefer most? Q4. Which of the two remaining options do you prefer least?
Pedersen (2016)	User orientation in general	1,379	4	3	First best; First worst; Second best;	Q1. Which of the three consultations would satisfy you most? Q2. Which of the three consultations would satisfy you least?
Van der Wulp (2012)	Health insurance coverage	2,000	5	4	First best; First worst	*Not detailed*
Xie (2014)	EQ-5D-5L states	100	8	3	First best; First worst	*Not detailed*
Yoo (2013)	Nursing jobs	526	8	3	First best; First worst	Q1. Which would you most like to get? Q2. Which would you least like to get?
Ghijben (2014)	Oral Anticoagulants	76	16	3	First best; Second best	Q1. Which would you choose? Q2. From the remaining two options, which would you choose?

Table 2 Attributes and levels used to describe personalisation of self-management programmes

Attributes	Levels *
Information (INFO)	(LOW) Provides everyone with the same information (HIGH) Provides information that is relevant to you
Situation (SITU)	(LOW) Takes little account of your current situation (HIGH) Makes suggestions that fit your current situation
Living well (LIVE)	(LOW) Seems to think that everyone wants to get the same from life (HIGH) Works with you on what you want to get from life
Communication (COMM)	(LOW) Communicates with you in a neutral professional way (HIGH) Communicates with you in a friendly and personal way
Cost per week (COST)	£5; £10; £15; £20

** LOW and HIGH refer to low and high levels of personalisation of the support for self-management (SSM) services*

Table 3 Descriptive analysis of personal characteristics in the experimental conditions

	WORST	BEST	BOTH
Sample size	**155**	**156**	**517**
Relationship (p1 = 0.1071; p2 = 0.5422)			
Not single	75.5	71.8	64.4
Single	24.5	28.2	35.6
Education level (p1 = 0.3984; p2 = 0.6862)			
Less than Univ.	56.1	53.2	57.4
Univ.	43.9	46.8	42.6
Employment (p1 = 0.0643; p2 = 0.8036)			
Not working	9.0	10.9	13.0
Retired	32.3	28.2	38.3
Disabled	14.8	17.3	13.5
Working	43.9	43.6	35.2
Health (p1 = 0.5484; p2 = 0.7537)			
Bad	37.4	39.7	43.7
Fair	24.5	26.3	22.4
Good	38.1	34.0	33.8
Gender (p1 = 0.0001; p2 = 0.3909)			
Male	56.8	51.3	34.0
Female	43.2	48.7	66.0
Annual income in £ (p1 = 0.0211; p2 = 0.5102)			
<= 15,599	12.9	19.2	28.2
[15,600-31,199]	33.5	30.8	34.0
>= 31,200	41.9	39.1	27.3
Not to say	11.6	10.9	10.4
Age in years (p1 = 0.0081; p2 = 0.0461)			
[18-40]	19.4	11.5	12.4
[41-50]	14.8	22.4	18.8
[51-60]	28.4	37.8	26.9
[61-70]	29.7	23.1	29.2
[71+]	7.7	5.1	12.8

p1: P-value of Chi-2 test comparing proportions between BEST and BOTH conditions
p2: P-value of Chi-2 test comparing proportions between BEST and WORST conditions

Table 4. Multinomial logit model allowing for an effect of type of choices on willingness-to-pay

	MLE	SE	P
1. Model parameters			
OPT2	0.100	0.046	0.029
OPT3	-0.020	0.043	0.645
COST	0.055	0.004	< 0.001
COST x [18-40] years	-0.012	0.008	0.111
COST x [41-50] years	0.021	0.007	0.004
COST x [61-70] years	0.000	0.006	0.969
COST x [71+] years	-0.015	0.011	0.168
Willingness-to-pay (WTP):			
INFO	8.582	1.221	< 0.001
SITU	15.307	1.485	< 0.001
LIVE	16.874	1.473	< 0.001
COMM	3.310	1.021	0.001
INFO x BEST	7.524	1.585	< 0.001
SITU x BEST	6.326	1.857	0.001
LIVE x BEST	1.527	1.569	0.331
COMM x BEST	-0.051	1.384	0.971
2. Model statistics			
# Individuals		311	
# Observations		3,732	
# Parameters		15	
Log-likelihood		-3,465.5	
BIC		7,054.4	

MLE: Maximum Likelihood Estimate; SE: Standard Error; P: P-value; BIC: Bayesian Information Criteria

Table 5. Heteroscedastic multinomial logit model allowing for an effect of type of choice (best versus worst) on errors variance

	MLE	SE	P
1. Preference parameters			
OPT2	0.097	0.045	0.031
OPT3	-0.029	0.042	0.486
INFO	0.699	0.040	< 0.001
SITU	1.021	0.052	< 0.001
LIVE	0.939	0.047	< 0.001
COMM	0.184	0.037	< 0.001
COST	-0.058	0.004	< 0.001
2. Scale parameters			
BEST	0.183	0.034	< 0.001
[18-40] years	-0.256	0.086	0.003
[41-50] years	-0.021	0.068	0.752
[61-70] years	0.128	0.058	0.028
[71+] years	0.186	0.097	0.056
3. Model statistics			
# Individuals		311	
# Observations		3,732	
# Parameters		12	
Log-likelihood		-3,464.4	
BIC		7,027.5	

MLE: Maximum Likelihood Estimate; SE: Standard Error; P: P-value; BIC: Bayesian Information Criteria

Table 6. Heteroscedastic multinomial logit model allowing for an effect of number of choices on errors variance

	MLE	SE	P
1. Preference parameters			
OPT2	0.181	0.031	< 0.001
OPT3	0.001	0.030	0.981
INFO	0.732	0.029	< 0.001
SITU	1.055	0.041	< 0.001
LIVE	0.867	0.035	< 0.001
COMM	0.243	0.026	< 0.001
COST	-0.056	0.003	< 0.001
2. Scale parameters			
BOTH	-0.156	0.024	< 0.001
Female	0.143	0.024	< 0.001
£[15600-31199]	0.016	0.034	0.634
£[31200+]	0.011	0.037	0.762
£["Not to say"]	0.006	0.051	0.911
[18-40] years	-0.222	0.059	< 0.001
[41-50] years	0.019	0.044	0.654
[61-70] years	0.134	0.038	< 0.001
[71+] years	0.102	0.054	0.061
3. Model statistics			
# Individuals		673	
# Observations		8,076	
# Parameters		16	
Log-likelihood		-7,488.8	
BIC		15,121.6	

MLE: Maximum Likelihood Estimate; SE: Standard Error; P: P-value; BIC: Bayesian Information Criteria

Table 7. Latent class logit model investigating the impact of number of choices on information processing strategies

	FIP			PIP "QUALITY"			PIP "COST"			NIP		
	MLE	**SE**	**P**	**MLE**	**SE**	**P**	**MLE**	**SE**	**P**	**MLE**	**SE**	**P**
1. Preference parameters												
OPT2	0.194	0.033	< 0.001	0.194	-	-	0.194	-	-	0.194	-	-
OPT3	-0.028	0.032	0.374	-0.028	-	-	-0.028	-	-	-0.028	-	-
INFO	1.166	0.041	< 0.001	1.166	-	-	0	-	-	0	-	-
SITU	1.816	0.065	< 0.001	1.816	-	-	0	-	-	0	-	-
LIVE	1.508	0.052	< 0.001	1.508	-	-	0	-	-	0	-	-
COMM	0.453	0.036	< 0.001	0.453	-	-	0	-	-	0	-	-
COST	-0.184	0.010	< 0.001	0	-	-	-0.184	-	-	0	-	-
2. Class membership parameters												
Constant	0	-	-	0.445	0.208	0.033	-0.711	0.238	0.003	-0.819	0.254	0.001
BOTH	0	-	-	0.494	0.171	0.004	0.596	0.207	0.004	0.827	0.222	< 0.001
Female	0	-	-	0.173	0.159	0.277	-0.385	0.181	0.034	-0.430	0.182	0.018
£[15600-31199]	0	-	-	-0.001	0.236	0.997	0.089	0.279	0.750	-0.019	0.293	0.949
£[31200+]	0	-	-	0.674	0.289	0.020	0.167	0.338	0.620	0.622	0.323	0.054
£["Not to say"]	0	-	-	-0.075	0.349	0.831	-0.075	0.421	0.859	-0.447	0.455	0.326
[18-40] years	0	-	-	-0.424	0.386	0.273	0.210	0.431	0.626	0.597	0.381	0.117
[41-50] years	0	-	-	0.189	0.306	0.536	-0.067	0.382	0.861	0.265	0.339	0.434
[61-70] years	0	-	-	0.169	0.264	0.522	-0.077	0.319	0.808	-0.560	0.337	0.097
[71+] years	0	-	-	-0.196	0.354	0.580	-0.670	0.450	0.137	-0.475	0.430	0.270
3. Model statistics												
# Individuals							673					
# Observations							8,076					
# Parameters							37					
Log-likelihood							-7,109.5					
BIC							14,551.8					

MLE: Maximum Likelihood Estimate; SE: Standard Error; P: P-value; BIC: Bayesian Information Criteria; FIP: full information processing; PIP: partial information processing; NIP: null information processing