Social Science & Medicine

To pay or not to pay? Cost information processing in the valuation of publicly funded healthcare

--Manuscript Draft--

Manuscript Number:	SSM-D-20-04179R2
Article Type:	Research paper
Keywords:	Multi-attribute choices; Cost attribute; Cost information processing; discrete choice experiment; Publicly funded healthcare
Corresponding Author:	Mesfin G Genie, PhD University of Aberdeen Aberdeen, Aberdeen UNITED KINGDOM
First Author:	Mesfin G Genie, PhD
Order of Authors:	Mesfin G Genie, PhD
	Mandy Ryan, PhD
	Nicolas Krucien, PhD
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Discrete choice experiments (DCEs) commonly include a monetary attribute. This enables willingness to pay (WTP), a monetary measure of benefit, to be estimated for non-monetary attributes. There has been concern that the inclusion of a cost attribute challenges the credibility of the experiment when valuing publicly funded healthcare systems. However, very little research has explored this issue. Using a UK sample, we allocated participants across two versions of a DCE: one including a cost attribute and the other excluding a cost attribute. The DCE was identical in all other respects. We find no significant difference in response time across the two surveys, monotonicity was higher for the COST DCE and cost was stated as the most commonly ignored attribute in the COST DCE. Whilst the inclusion of a cost attribute did not alter the structure of preferences, it resulted in a lower level of choice consistency. Using an unrestricted latent class model, we find evidence of a credibility effect: respondents with experience of paying for health services and who perceive the choices as realistic are less likely to ignore cost. Further, respondents with a higher response time are less likely to be cost minimisers. Results are robust across different model specifications and choice formats. DCE practitioners should give due consideration to cost credibility when including a cost attribute, ensuring participants engage with the cost attribute. Ways to do this are suggested, including careful motivation of the cost attribute, consideration to the appropriate payment vehicle and careful consideration to the cost attribute when developing and piloting the survey. Failure to do this will result in an invalid willingness to pay estimates and thus policy recommendations.

Title page

To pay or not to pay? Cost information processing in the valuation of publicly funded healthcare

Mesfin G. Genie[a]
Mandy Ryan[a]
Nicolas Krucien[b]

[a]Health Economics Research Unit, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom, Email: m.ryan@abdn.ac.uk
[b]Patient-Centered Research, Evidera, London W6 8BJ, United Kingdom, Email: Nicolas.Krucien@evidera.com

Corresponding author
Dr Mesfin G. Genie
Health Economics Research Unit
University of Aberdeen
Polwarth Building
Foresterhill
Aberdeen, AB25 2ZD, United Kingdom
Email: mesfin.genie@abdn.ac.uk

Acknowledgments
We thank all participants who took part in the study and three anonymous reviewers for their comments on the paper. The design of the discrete choice experiment on which this paper draws was shaped by a team that included, alongside two of the authors, Professor Chris Burton, Professor Vikki Entwistle, Professor Christine Bond, Dr Terry Porteous and Dr Alison Elliott.

Conflict of interest
The authors have no conflict of interest to declare whatsoever.

Funding sources
The original DCE study was funded by the Health Foundation. This work has been funded by the University of Aberdeen and the Chief Scientist Office of the Scottish Government Health and Social Care Directorates. These sponsors were not involved in the study design, data analysis and writing of the article. The information and views set out in the article are those of the authors.

ABSTRACT

Discrete choice experiments (DCEs) commonly include a monetary attribute. This enables willingness to pay (WTP), a monetary measure of benefit, to be estimated for non-monetary attributes. There has been concern that the inclusion of a cost attribute challenges the credibility of the experiment when valuing publicly funded healthcare systems. However, very little research has explored this issue. Using a UK sample, we allocated participants across two versions of a DCE: one including a cost attribute and the other excluding a cost attribute. The DCE was identical in all other respects. We find no significant difference in response time across the two surveys, monotonicity was higher for the COST DCE and cost was stated as the most commonly ignored attribute in the COST DCE. Whilst the inclusion of a cost attribute did not alter the structure of preferences, it resulted in a lower level of choice consistency. Using an unrestricted latent class model, we find evidence of a credibility effect: respondents with experience of paying for health services and who perceive the choices as realistic are less likely to ignore cost. Further, respondents with a higher response time are less likely to be cost minimisers. Results are robust across different model specifications and choice formats. DCE practitioners should give due consideration to cost credibility when including a cost attribute, ensuring participants engage with the cost attribute. Ways to do this are suggested, including careful motivation of the cost attribute, consideration to the appropriate payment vehicle and careful consideration to the cost attribute when developing and piloting the survey. Failure to do this will result in an invalid willingness to pay estimates and thus policy recommendations.

Keywords: Multi-attribute choices; Cost attribute; Cost information processing; Discrete choice experiment; Publicly funded healthcare

JEL codes : D01, D80, C35, D90, I12

To pay or not to pay? Cost information processing in the valuation of publicly funded healthcare

1. Introduction

Discrete choice experiments (DCEs) are commonly used in health economics to investigate individuals' preferences for multi-attribute services (Clark et al., 2014; de Bekker-Grob et al., 2012; Soekhai et al., 2019). DCEs are appealing given they are grounded in microeconomic theory (Lancaster, 1966; Manski, 1977), thus allowing welfare measures to inform policy decisions. When a monetary attribute is included in the DCE (e.g., out-of-pocket expense for medical services), willingness-to-pay (WTP) for changes in services can be estimated. These WTP values can be used within a cost-benefit analysis to inform health policy (McIntosh, 2006).

DCEs have been criticised for their hypothetical nature, with questions raised about the validity of welfare estimates generated (Carson et al., 2014; Vossler et al., 2012; Zawojska & Czajkowski, 2017). Ex-ante correction procedures, such as "cheap-talk scripts" (Broadbent, 2014; Carlsson et al., 2011, 2013; Fifer et al., 2014; Lusk, 2003; Silva et al., 2012) and "oath protocols" (De-Magistris & Pascucci, 2014; Jacquemet et al., 2013; Kemper et al., 2016; Stevens et al., 2013) have been developed to address such concerns. However, such procedures ignore the credibility of the cost attribute in the context of publicly provided goods (Carson & Groves, 2007). Zawojska et al. (2019) note that perceived payment consequentiality is related to the credibility of the cost attribute. Lack of credibility may lead participants to change their choice behaviour, for example, by ignoring the cost attribute (Pedersen et al., 2011; Ratcliffe, 2000; Sever et al., 2019).

Using split sample designs, four studies have investigated the effect of including a cost attribute on stated choices within a DCE (Bryan et al., 1998; Essers et al., 2010; Pedersen et al., 2011; Sever et al., 2019). All provide evidence that inclusion of the cost attribute does not affect preference ranking for non-monetary attributes (though Pedersen et al. (2011) found evidence of an effect in forced choices). Three of the studies took place in a publicly provided healthcare system where individuals are not used to paying for health care at the point of consumption. In such a context, the inclusion of the cost attribute may question the credibility of the DCE. Bryan et al. (1998), in a UK study looking at preferences for the diagnosis and treatment of severe knee injuries, and using a forced-choice DCE, found that including the cost attribute did not affect the structure of preferences for non-monetary attributes. However, the cost attribute was not significant. Using an unforced choice, Essers et al. (2010) explored preferences for surgical treatment of primary basal cell carcinoma in the Netherlands. Whilst the cost attribute was statistically significant, its inclusion did not affect the structure of preferences for non-monetary attributes. Pedersen et al. (2011) investigated patients' preferences for organisational characteristics of general health care practice in Denmark. Whilst inclusion of the cost attribute did not affect the structure of preferences or the error variance in an unforced choice, in a forced choice DCE, the cost attribute changed the structure of preferences and lowered choice consistency. More recently, in a study investigating preferences for the delivery of dental care in Croatia, Sever et al. (2018) found that whilst the inclusion of the cost attribute did not affect the structure of preferences, it increased the response error variance.

However, these studies did not consider decision heuristics, which would mask issues with the cost attribute. For example, if some respondents only consider the cost attribute due to complexity (i.e. large cost coefficient) and others ignore cost due to lack of credibility (i.e., small coefficient), the average preference would mask such effects. Further, Pedersen et al.

(2011) and Sever et al. (2019) found a negative effect of cost inclusion on choice consistency (or response error variance). These findings may be explained by an increase in the complexity of the choice tasks and/or a decrease in the perceived cost credibility. Notably, Sever et al. (2019) conducted their study in Croatia, where the public is familiar with paying for care. This raises the question of whether the increased error variance is a result of complexity rather than credibility.

Using a split sample design and following previous literature, we first investigate whether including a cost attribute impacts preference rankings and choice consistency. We then provide new evidence on the effect of payment experience (a proxy for cost attribute credibility) on how respondents process the cost attribute. The remainder of the paper is organised as follows. Section 2 describes the study context, the design of the experiment, and the comparison of the two DCEs based on three indicators (monotonicity, response times and stated attribute non-attendance). In Section 3, we compare preference rankings and choice consistency. Section 4 investigates cost information processing strategies, with attention to the role of credibility. Section 5 discusses the implications of our results for DCEs conducted within publicly funded healthcare systems and identifies avenues for future research. Section 6 draws conclusions.

2. Discrete choice experiment

2.1. Context

The context is patient preferences for personalisation of chronic pain self-management (CPSM) programmes in the UK. Detailed information on the study is available in (Burton et al., 2017). Briefly, we used focus groups to identify attributes and levels. In addition to the cost attribute ("COST"), four qualitative attributes were included (Table 1): providing personalised information ("INFORMATION"); providing advice that matches personal situation ("SITUATION"); putting an emphasis on personal values in living well ("LIVE WELL"); and communication style ("COMMUNICATION"). These attributes were developed from prior extensive quantitative and qualitative work and refined and tested through a series of stages following best practices for DCE development (See details in **Online Supplementary Material** (**OSM-A**)). The cost attribute levels were derived from the quantitative pilot study using the direct willingness to pay (WTP) approach (see **OSM-B**).

Our qualitative survey development work, using think-aloud interviews, suggested challenges with the cost attribute (Burton et al., 2017). We found evidence of cost-based responses (*I'd rather pay ten, fifteen, twenty pounds and have face-to-face contact than five pounds, knowing that that would have to be either a large group, seminar-type thing or an online thing, I'd rather pay more and have a face-to-face contact);* cost attribute non-attendance (*I didn't take any account of the cost, because if it was going to help me, I would pay for it)* and protest responses (*I think, morally, there shouldn't be that question, it should be free. […] I think if you've got a long-term condition you should be entitled to some support).* The quantitative pilot study also found evidence of protests (*"I think that the NHS should pay for the bill as I have paid in all of my working life and now that I need to use the service, I consider that my payments should cover the cost", "I have worked all of my adult life and paid taxes and national insurance, so I deserve it. Unlike those that get the same but have never contributed, my situation is that I do not need any support at the moment if I did, then I could only pay a minimal amount".*

Concerns about the inclusion of the cost attribute within the research team led to a NOCOST DCE to investigate the influence on respondents' choice behaviour. The follow-up "NOCOST

DCE" sample was recruited two months after the main study (COST DCE). This questionnaire was identical to the COST DCE in all other respects.

2.2. Experimental design

A main effects efficient design was employed to devise the choice tasks (Rose & Bliemer, 2013), resulting in 12 experimental choice tasks in both the COST and NOCOST DCE (each including three unlabelled choice options). Two non-experimental tasks were added: Task #1 was a 'warm-up' choice with the aim of familiarising respondents with the choice format and Task #14 was a monotonicity test (i.e., do respondents choose dominant options). DCEs are commonly designed to maximise the level of precision for a given sample size (referred to as D-efficient design). However, it is possible to follow an alternative approach, minimising the sample size required for a given precision level (referred to as S-efficient design) (Louviere et al., 2008). Our design followed a two-stage process: a quantitative pilot DCE (n = 120) using a D-efficient design with non-informative priors optimised for a multinomial logit (MNL) model, and the main DCE using an S-efficient design optimised for the estimation of MNL model with fully informative priors obtained from the quantitative pilot study. The experimental design procedure was the same in both conditions; for the NOCOST DCE, we removed the cost attribute. This minimises biases from different experimental designs.

Given the study aimed to estimate personal preferences, we used the Best Worst (BW) DCE approach. It has been argued that the extra information generated from the BW DCE approach enables individual-level preferences to be estimated (Lancsar et al., 2013). Participants were asked to choose both their most and least preferred options: they first selected their least preferred option and then selected their most preferred option. For each choice, they were presented with a follow-up question, asking if they would buy the best option (see Figure 1).

Veldwijk et al. (2014) argued that since respondents learn from answering forced-choice tasks, a dual response design might result in higher data quality than offering a direct opt-out option. Using the same data set used in our study, Krucien et al. (2019) found that the determinants of choice behaviour differ between these types of decisions; therefore, following Lancsar et al. (2013), we only considered the most preferred choices alongside the purchase decision.

2.3. Recruitment and sample

We worked with a market research company, ResearchNow! (now called Dynata), to recruit 500 respondents for the COST DCE (517 achieved) and 200 for the NOCOST DCE (206 achieved). (See **OSM-C** for the details of minimum sample size calculation.) 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 3 months. The study was approved by the North of Scotland Research Ethics Service (Reference 14/NS/0075).

The DCE questionnaire collected information on respondents' characteristics and experience of paying for CPSM services (Table 2). The two samples significantly differed only in terms of gender, with more women responding to the COST DCE (66%) than the NOCOST DCE (52%) (Chi-2=12.28; $p<0.001$).

2.4. Monotonicity, Response Time and Stated Attribute Non-Attendance

In addition to the choice tasks, which included a monotonicity test, we collected data on the time respondents took to complete each choice task and stated attribute non-attendance (SANA). We compared these across the COST and NOCOST DCEs.

Mattmann et al. (2019) argued that non-monotonicity is likely in DCEs when respondents are presented with unfamiliar attributes (e.g. cost). However, the opposite may also occur with higher monotonicity in the COST DCE where it is easier to identify the dominant alternative. This is especially the case given the other attributes in our study are qualitative and therefore not entirely monotone, e.g. some respondents may prefer communication in a neutral and professional way over communication in a friendly and personal way. (We thank an anonymous referee for this point.) We found significantly more respondents passed the monotonicity test in the COST DCE (86.5%) compared to the NOCOST (79.1%) (Chi-square = 6.008, p=0.014).

We measured response time (RT) in seconds at the choice task level, measured as the time between the display of the task on the screen and when individuals clicked the "next task" button. We computed the average RT that respondents took to complete the 12 choice tasks (i.e., the average response time per respondent). Fast response time could indicate either a "clicking-through" choice behaviour or participants being highly efficient in making choices. It can also be argued that respondents who take a longer time to complete the choice tasks are expected to engage in more cognitive reasoning (Dellaert et al., 2012). We found no significant differences in RT (COST DCE: mean RT = 245.495 seconds, SD=965.589 seconds; NOCOST DCE: mean RT = 234.78 seconds, SD=400.3 seconds; t=1.067, p=0. 286).

After completing the choice tasks, respondents were asked how often they considered the different attributes (Table 3). Responses to these SANA questions did not differ statistically between the two DCEs. However, for the COST DCE group, COST was more frequently '*never considered*' (10% for COST compared to 3% for INFORMATION, 2.9% for SITUATION, 3.5% for LIVING WELL and 6% for COMMUNICATION).

3. Comparison of preference ranking and choice consistency

3.1. Methods

We first compared relative importance (RI) of attributes across the COST and NOCOST DCEs. The analysis of DCE responses is based on the random utility model (RUM) (McFadden, 1974), which stipulates that the utility (U) of the choice option (j) faced by the respondent (n) in a choice task (t) depends on a systematic component (V) which can be explained by the included attributes and a stochastic component (ε) which is unobservable.

$$U_{ntj} = \lambda_{ntj} V_{ntj} + \varepsilon_{ntj} \tag{1}$$

The stochastic component is typically assumed to be identically and independently distributed as type 1 extreme value ($\varepsilon_{ntj} \sim iidEV1$) (McFadden, 1974), which leads to the multinomial logit model (MNL). λ_{ntj} is a scale parameter inversely related to the variance of the stochastic component (σ_{ε}). Since the scale parameter is perfectly confounded with parameters of preferences, it is usually assumed to be equal to one for identification purposes, which makes σ_{ε} a fixed quantity (Train, 2009). The indirect utility function (V) is typically assumed to be linear and additive and can be expressed as:

COST DCE: $V^{COST} = \beta_{11} ASC_B + \beta_{12} ASC_C + \beta_{13} ASC_{NO\_BUY} + \beta_{14} INFO + \beta_{15} SITU + \beta_{16} LIVE + \beta_{17} COMM + \beta_{18} COST$ (2)

NOCOST DCE: $V^{NOCOST} = \beta_{21}ASC_B + \beta_{22}ASC_C + \beta_{23}ASC_{NO\_BUY} + \beta_{24}INFO + \beta_{25}SITU + \beta_{26}LIVE + \beta_{27}COMM$ (3)

where $\beta_{14:18}$ and $\beta_{24:27}$ are the estimated preference parameters capturing the marginal sensitivity to changes in the attributes; ASC_B and ASC_C are alternative specific constants showing a general preference to choose Option B and C over A with $\beta_{11:12}$ and $\beta_{21:22}$ the corresponding coefficients; ASC_{NO_BUY} is the alternative specific constant showing a general preference to opt-out (and not buy the option) with the corresponding preference parameters (β_{13} and β_{23}). All labels are defined in Table 1.

After estimating a multinomial logit (MNL) model for each DCE, we obtained attribute relative importance (RI) score using Equation 4. We used a bootstrapping procedure with 1,000 repetitions to obtain a 95% confidence interval around the RI scores (Orme, 2010):

$$RI_k = \frac{Max(PWU_k) - Min(PWU_k)}{\sum_k (Max(PWU_k) - Min(PWU_k))} \qquad (4)$$

where PWU_k corresponds to the part-worth utility (coefficient) of the k[th] attribute. For example, the costs levels range from £5 to £20; for a cost sensitivity of -0.088, the minimum PWU is 5 x (-0.088) = 0.44 and the maximum PWU is 20 x (-0.088) = 1.76.

We then allow the error variance to systematically differ between the COST and NOCOST DCE samples. Changes in error variance represent differences in choice consistency (Börger, 2016; DeShazo & Fermo, 2002). Following previous literature, we first estimated a heteroscedastic version of the MNL (HMNL) model on the pooled data (Equation 5), with error variance a function of DCE type (COST or NOCOST) (Hole, 2006; Swait & Adamowicz, 2001). We then estimated a heteroscedastic mixed logit (HMXL) model (McFadden & Train, 2000; Train, 2009), accounting for the panel structure of the data and relaxing the IIA assumption whilst modelling preference heterogeneity and scale differences between the COST and NO COST DCEs (Börger, 2016; Czajkowski et al., 2016). Cost was assumed to be log-normally distributed to restrict the sign to be the same for all respondents. The model was estimated using the maximum simulated likelihood method with 1000 Sobol draws (Czajkowski & Budziński, 2019). We expect less consistent choices in the COST DCE ($\alpha_1 < 0$). Given the two samples differ in terms of gender, we also measured the effect of gender on the scale. By including gender in the scale function, we obtain a "cost effect" controlling gender differences.

$$V^{POOLED} = \lambda_n(\beta_1 ASC_B + \beta_2 ASC_C + \beta_3 ASC_{NO\_BUY} + \beta_4 INFO + \beta_5 SITU + \beta_6 LIVE + \beta_7 COMM + \beta_8 COST) \quad (5)$$

$$\lambda_n = \exp(\alpha_1 COSTDCE + \alpha_2 FEMALE)$$

3.2. Results

The MNL results are presented in columns 1 and 2, Table 4. All utility coefficients have the expected signs, confirming the theoretical validity of the model. The RI scores are presented in Figure 2. For the COST DCE, the cost attribute was the second most important attribute with a RI score of 22.9% [19.8%; 26.1%]. SITUATION was the most important attribute with a score of 27.7% [25.7%; 29.7%], and COMMUNICATION the least important with a score of 7.2% [5.6%; 8.9%]. In the NOCOST DCE, SITUATION was the most important attribute with a RI score of 37.02% [34.2%; 40.1%], and COMMUNICATION was the least important with a RI score of 8.97 [6.1%; 11.8%]. The ranking of non-monetary attributes did not differ significantly across the two DCEs (Spearman correlation=1).

Results of the heteroscedastic MNL and MXL models are shown in columns 3 and 4, Table 4. As expected, α_1 was negative and significant, indicating less consistent choices in the COST DCE.

Whilst previous studies attributed increased response error variance to increases in choice task complexity (Dellaert et al., 2012) and handling a larger number of attributes in a choice set (DeShazo & Fermo, 2002; Islam et al., 2007), cost credibility may also make people less certain about their preferences. This may encourage respondents to simplify the choice task by ignoring cost or considering only cost, thereby increasing the response error variance. Although Pedersen et al. (2011) asked respondents how difficult they perceived the choice tasks, it is unclear whether the higher error variance in the forced choice was due to perceived difficulty. Further, Bryan et al. (1998) investigated the number of missing observations (which could be considered a proxy for complexity) between the two DCEs, but its effect on cost information processing and choice consistency is unclear. Flores & Strong (2007) explored the implications of costs not perceived to be credible by survey participants in stated preference analysis of public goods and noted that less credibility implies a larger variance. Using the COST DCE, we next explore cost processing strategies with attention to the role of credibility.

4. Understanding cost information processing strategies

Respondents might give an unequal amount of attention to attributes used in a DCE: they might pay more attention to attributes considered to be more important and much less attention to or even ignore attributes that are considered to be less important (Koetse, 2017). The reasons for this behaviour may vary from time pressure to cognitive overload/task complexity (Kardes et al., 2004; Lussier & Olshavsky, 1979) to attribute credibility (Hensher, 2007; Sælensminde, 2006). Such heuristics may be particularly pronounced for the cost attribute in the case of public goods (e.g., cost attribute may not be credible when people are not familiar with payment for health care services in a publicly funded healthcare system).

4.1. Methods

We explore the decision-making strategies/pattern of preferences employing an unrestricted latent class model (hereafter LCM) to produce a set of classes, each representing a pattern of valuation of the cost attribute. The choice probability that a respondent n of class q chooses alternative i from a particular set J, comprising j alternatives, is expressed as (Scarpa et al., 2013; Shen, 2009):

$$P_{n|q} = \prod_t \frac{\exp(\sum_k \beta_{qk} X_{ntik})}{\sum_{j=1}^{J} \exp(\sum_k \beta_{qk} X_{ntjk}))} \qquad q = 1, \dots, Q \tag{6}$$

where β_{qk} are the average preferences for attribute k in latent class q associated with the vector of explanatory variables X_{ntik}, t corresponds to the choice tasks, n to the respondents, k to the attributes, j to the choice options, and i denotes the chosen option. $P_{n|q}$ is the probability of all the choices made by individual n conditional on being in class q.

The underlying theory of the LCMs suggests that respondents' choice behaviour and preferences are allocated into a set of Q latent classes. Preferences within each class are assumed to be homogenous but allowed to differ across classes. The LCM estimates Equation 6 for Q classes and predicts the probability H_{qn} as respondent n being in class q. Then, the probability of individual n belonging to class q, H_{qn}, is given as:

$$H_{qn} = \frac{\exp(\sum_s \alpha_{qs} Z_{ns})}{\sum_q \exp(\sum_s \alpha_{qs} Z_{ns})} \tag{7}$$

where s denotes the personal characteristics (e.g., income, payment experience, etc.), Z_{ns} is the value of the s^{th} characteristic for respondent n. α_{qs} capture the effects of the personal characteristics on the class membership.

The unconditional probability of the choices made by individual n, P_n, is given as:

$$P_n = \sum_q H_{qn} P_{n|q} \tag{8}$$

We allocated individuals across classes by combining Bayes theorem with the maximum probability allocation rule (Greene & Hensher, 2003); the class share represents the proportion/percentage of respondents belonging to each class.

$$\hat{H}_{q|n} = \frac{\hat{P}_{n|q} \hat{H}_{qn}}{\sum_q \hat{P}_{n|q} \hat{H}_{qn}} \tag{9}$$

Explanatory variables of class membership and preference parameters of respondents in each class are estimated jointly. We included the following covariates of class membership:

EXPERIENCE captures the respondent's experience of paying for health care (a proxy for cost credibility). We asked respondents if they had any payment experience of CPSM services during the past six months (e.g. pain management programme; pain clinic; physiotherapist; other therapists (acupuncture, osteopath, and chiropractor); or other CPSM services) (Table 2). We hypothesise that respondents with experience of paying for CPSM services would be more likely to consider the cost attribute as credible and then less likely to adopt cost processing heuristics than those who had no experience of paying.

RESPONSE TIME (RT) captures the average time respondents took to complete the 12 choice tasks and is a proxy for the ability to make choices. Following previous studies (Börger, 2016; Börjesson & Fosgerau, 2014; Campbell et al., 2018), we used response time (RT) to approximate respondents' ability to make choices. Börger (2016) found that response time for the choice tasks decreased error variance and indicated that respondents who report stronger attribute attendance take longer to complete the choice tasks. Consistent with Börger (2016), we also find error variance lower among respondents who took a longer time to complete the choice task for both forced and unforced data (results are available on request from the authors). However, Campbell et al. (2018) find the opposite result: error variance is highest among those who took the longest time to complete the choice experiment. Respondents who find making choices easier are expected to make choices more quickly. However, it can also be argued that respondents who take a longer time to complete the choice tasks are expected to engage in more cognitive reasoning.

SERVICE IN MIND. It is argued that information about a good (e.g., obtained from expert advice) can affect decision-making (Eil & Rao, 2011; Grossman & Owens, 2012; Schotter, 2003). We asked individuals whether they had a particular type of CPSM service in mind when completing the choice tasks. This information was converted into a binary variable indicating whether they had a service in mind (*Either online, group, one-to-one, or other types of support*) or not (*Either no service in mind or not sure*) (see Table 2). This information was included as a proxy for REALISM: we hypothesise that respondents with a service in mind would perceive the choice as more realistic and then less likely to adopt cost-processing heuristics than those who had no service in mind.

The optimal number of latent classes is a trade-off between explanatory power, the number of additional parameters, and ease of interpretation (Czajkowski et al., 2020). Further, it has been suggested that the statistical criteria and the significance of the parameter estimates need to be tempered by the analyst's own judgement of the suitability of the model (Hynes et al., 2008; Scarpa & Thiene, 2005). We estimated five latent class models, including between two and six classes and retained the model with the lowest Bayesian Information Criterion (BIC) (Czajkowski, 2020; Greene & Hensher, 2003; Shen, 2009).

4.2. Results

The unconstrained LCM results are presented in Table 5: the upper part shows the coefficients/preferences of the attributes for each class, and the lower part indicates the effect of covariates on class membership.

The optimal number of classes was six, differing in terms of preference parameters. Since our interest is in understanding cost preferences, our interpretation focuses on the cost attribute and the 'purchase/no purchase' decision. Classes 1 (8.6% of respondents) and 2 (46.5% of respondents) correspond to a strong preference for a 'no purchase' decision (i.e., those who would not buy the option they like the most). The insignificant cost coefficient suggests that respondents more likely to be in these classes are more likely to ignore the cost attribute; we label such respondents "**cost ignorers**". Parameters of Class 3 (13.7%), Class 4 (12.6%), and Class 6 (12.6%) indicate that respondents with a high probability of belonging to these classes are more likely to consider cost alongside other attributes, although cost sensitivities differ across classes; we label such respondents "**cost compensators**". Latent Class 5 (6%) comprises respondents who are less likely to make a 'no-purchase' decision and more likely motivated by cost minimisation considerations; we label such respondents "**cost minimisers**".

In looking at characteristics that predict class membership, EXPERIENCE of payment has a negative and statistically significant effect on the probability of belonging to Classes 1 and 2. That is, respondents who have experience paying for CPSM services (i.e., who perceive the cost as credible) are less likely to be cost ignorers and to make a 'no purchase' decision. This finding supports our credibility hypothesis. Further, respondents who had a particular type of support service in mind (which can be viewed as a proxy for perceived realism) are less likely to belong to be class 2; they are less likely to ignore cost and less likely to make a 'no purchase' decision. Respondents with longer RT are less likely to be cost minimisers (Class 5).

5. Discussion

DCEs typically include a monetary attribute to allow derivation of WTP measures. When individuals have no experience of paying for health care, the credibility of the DCE may be questioned. We find no significant difference in response time between the COST and NOCOST DCEs, monotonicity is higher in the COST DCE, and respondents state they ignored the cost attribute and stated attribute non-attendance was not affected for other attributes. Consistent with existing literature, we found that the inclusion of the cost attribute did not influence the preference ranking of attributes but did result in less consistent responses. Exploring cost information processing strategies, we identify three groups: **cost-ignorers**, **cost-minimisers** and **cost-compensators**. Class membership results suggest credibility effects: respondents who had experience of paying for CPSM services are less likely to ignore cost. Further, respondents with a higher response time are less likely to adopt cost-minimisation. This may be explained by increased engagement.

Our results have practical implications for DCE practitioners. If participants ignore the cost attribute or focus on the cost attribute as a signal to simplify their choices (e.g., cost minimisation or cost ignorance), the validity of WTP estimates is questioned. This is confirmed by previous studies showing considering attribute non-attendance (ANA) impacts WTP estimates (Carlsson et al., 2010; Hensher & Greene, 2010). Therefore, such decision heuristics should be explored analytically before using the DCE results for policy.

Consistent with the existing literature (Hensher et al., 2012; Hess & Hensher, 2010), we found an inconsistency between inferred ANA and stated ANA. Both inferred ANA and Stated ANA (SANA) have their limitations. It is not clear whether econometric models inferring ANA are reliable. The use of econometric modelling to explore ANA confounds ANA as a decision rule with true preferences (Hess et al., 2013). For example, the "cost ignorance" rule can be a decision rule *per se* (i.e., a decision to ignore cost systematically) or a case where people have a little true sensitivity to price changes such that cost is not ignored but does not really matter (at least not in the range of cost included in the experiment). Further, SANA is not exempt from limitations, including misinterpretation of the SANA questions with results depending on the format of SANA questions (van Loo et al., 2018); inability to recall (Caputo et al., 2018), and untruthful responses (Kragt, 2013). As noted by Heidenreich et al. (2018), given the limitations of econometric modelling and subjective reflections, think-aloud interview methods (Ryan et al., 2009) and eye-tracking (Balcombe et al., 2015) may provide more insights into causes of ANA in different applications, especially at the design and evaluation stage of DCE studies.

Our results suggest DCE practitioners should give careful consideration to the inclusion of the cost attribute in a DCE. Including 'cheap talk' (script or voice clip) to encourage more attentive valuation of cost attribute may increase its credibility (Özdemir et al., 2009). Another strategy to increase credibility is to use alternative payment vehicles. Gafni (1991) highlighted the importance of using payment vehicles that resemble reality. Posavac (1998) indicated that an inappropriate choice of payment vehicle might lead to hypothetical bias. Smith (2003) noted that the most suitable payment format will depend upon the study context and differ across cultures, countries and products.

Limited guidance is provided in health care DCE studies on how the payment vehicle (cost attribute) is defined, the wording of the cost attribute, and the forms and frequency of payment. Most published studies either do not report how the cost attribute is described or are extremely brief in their description of the cost attribute. For instance, previous studies in the context of diabetes worded the cost attribute in different ways: "cost of diabetes medicines each month" (Johnson et al., 2011), "personal cost to you each month" (Mohamed et al., 2013), and "payment per month out of pocket" (Feher et al., 2016). Although the wording and form of the payment vehicle are context-dependent, we suggest DCE practitioners take time when constructing the cost attribute in terms of the wording, format, and frequency. It would be particularly interesting to explore how the wording of the cost attribute, frequency (e.g. monthly versus annually), and types of payment vehicle in publicly funded health care system (e.g. charity donation, direct out-of-pocket expenditure, changes in medical insurance premiums) affect choice behaviour in health DCEs; this will allow us to learn more on the cost attribute and to identify best-practice methods for incorporating cost in health care DCEs. We leave this as a further area of research.

Our study raises a number of issues and suggests areas for future research. We did not randomly allocate respondents across the two experimental conditions. When comparing socio-demographic characteristics of participants across the samples, only gender was significantly different. However, as a robustness check, we draw a stratified sample by gender of 206 respondents from the COST DCE of 517 respondents (we thank the anonymous referee for this suggestion). Consistent with our initial results, including cost did not affect the ranking of preferences for the non-monetary attributes. While the effect of gender on the error variance disappears, including the cost attribute still increased the error variance (See results in **OSM-D**).

In terms of the context and generalisability of results, the best worst (BW DCE) format may not be typical of many DCEs, although the method is being increasingly used (Krucien et al., 2019). Given that the BW DCE format is argued to be easier for respondents to answer (Lancsar et al., 2013), it might be argued that simple decision-making heuristics are less likely, and then our findings are important, i.e., even for a simpler approach, there is evidence of decision-making heuristics. Further, from a data quality point of view, a dual-response choice format (Brazell et al., 2006) has been shown to be less open to ambiguity than the opt-out alternative from the traditional DCE (Carlsson et al., 2007) and mimic actual choices in a market situation better (Ryan & Skåtun, 2004). We suggest future research explores the inclusion of a cost attribute using other DCE formats. We note here that our results are robust for a forced choice data (i.e., without the "purchase/no-purchase" decision); see **OSM-E).**

Our study focused on the effects of including a cost attribute in a health care context where people have limited experience of paying for NHS care; however, a private market for the management of chronic pain does exist (e.g., physiotherapy). In other DCE applications, such as preferences for new cancer treatment, a cost attribute may be 'more' unrealistic, and therefore its inclusion in the DCE becomes more problematic. In developing countries, where there is limited ability to pay, the cost attribute may be more challenging. It has been suggested that the payment vehicles used in developed countries should be reconsidered for suitability when conducting DCEs in developing country contexts (Gibson et al., 2016; Hassan et al., 2018). Future research should explore the impact of the cost attribute on the structure of preferences in different health care contexts and different country settings.

6. Conclusion

Monetary valuation of health care is problematic when respondents are not used to paying out-of-pocket fees for health services; they are more likely to ignore cost or consider only cost. Almost 61% of our respondents did not trade the cost attribute against other features. This has implications for the validity of WTP measures derived from DCEs and their use in economic evaluation/policymaking. Consideration should be given to ways to engage respondents with the cost attribute, thus increasing cost credibility and resulting welfare estimates.

References

Balcombe, K., Fraser, I., & Mcsorley, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*, *30*(3). https://doi.org/10.1002/jae.2383

Börger, T. (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). https://doi.org/10.1007/s10640-015-9905-1

Börjesson, M., & Fosgerau, M. (2014). Response time patterns in a stated choice experiment. *Journal of Choice Modelling*, *14*. https://doi.org/10.1016/j.jocm.2015.02.001

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P. F. (2006). The no-choice option and dual response choice designs. *Marketing Letters*, *17*(4). https://doi.org/10.1007/s11002-006-7943-8

Broadbent, C. D. (2014). Evaluating mitigation and calibration techniques for hypothetical bias in choice experiments. *Journal of Environmental Planning and Management*, *57*(12). https://doi.org/10.1080/09640568.2013.839447

Bryan, S., Buxton, M., Sheldon, R., & Grant, A. (1998). Magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences. *Health Economics*, *7*(7). https://doi.org/10.1002/(SICI)1099-1050(1998110)7:7<595::AID-HEC381>3.0.CO;2-E

Burton, C. D., Entwistle, V. A., Elliott, A. M., Krucien, N., Porteous, T., & Ryan, M. (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). https://doi.org/10.1136/bmjopen-2016-015689

Campbell, D., Mørkbak, M. R., & Olsen, S. B. (2018). The link between response time and preference, variance and processing heterogeneity in stated choice experiments. *Journal of Environmental Economics and Management*, *88*. https://doi.org/10.1016/j.jeem.2017.10.003

Caputo, V., van Loo, E. J., Scarpa, R., Nayga, R. M., & Verbeke, W. (2018). Comparing Serial, and Choice Task Stated and Inferred Attribute Non-Attendance Methods in Food Choice Experiments. *Journal of Agricultural Economics*, *69*(1). https://doi.org/10.1111/1477-9552.12246

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2007). Consumer willingness to pay for farm animal welfare: Mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics*, *34*(3). https://doi.org/10.1093/erae/jbm025

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Löfgren, åsa, Qin, P., & Sterner, T. (2013). The truth, the whole truth, and nothing but the truth-A multiple country test of an oath script. *Journal of Economic Behavior and Organization*, *89*. https://doi.org/10.1016/j.jebo.2013.02.003

Carlsson, F., Kataria, M., & Lampi, E. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental and Resource Economics*, *47*(1). https://doi.org/10.1007/s10640-010-9365-6

Carlsson, F., Martinsson, P., & Akay, A. (2011). The effect of power outages and cheap talk on willingness to pay to reduce outages. *Energy Economics*, *33*(5). https://doi.org/10.1016/j.eneco.2011.01.004

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*(1). https://doi.org/10.1007/s10640-007-9124-5

Carson, R. T., Groves, T., & List, J. A. (2014). Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2). https://doi.org/10.1086/676450

Clark, M. D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, E. W. (2014). Discrete Choice Experiments in Health Economics: A Review of the Literature. In *PharmacoEconomics* (Vol. 32, Issue 9). https://doi.org/10.1007/s40273-014-0170-x

Czajkowski, M., & Budziński, W. (2019). Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling*, *31*. https://doi.org/10.1016/j.jocm.2019.04.003

Czajkowski, M., Gajderowicz, T., Giergiczny, M., Grotkowska, G., & Sztandar-Sztanderska, U. (2020). Choosing the Future: Economic Preferences for Higher Education Using Discrete Choice Experiment Method. *Research in Higher Education*, *61*(4). https://doi.org/10.1007/s11162-019-09572-w

Czajkowski, M., Hanley, N., & LaRiviere, J. (2016). Controlling for the Effects of Information in a Public Goods Discrete Choice Model. *Environmental and Resource Economics*, *63*(3). https://doi.org/10.1007/s10640-014-9847-z

de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics*, *21*(2). https://doi.org/10.1002/hec.1697

Dellaert, B. G. C., Donkers, B., & van Soest, A. (2012). Complexity effects in choice experiment-based models. *Journal of Marketing Research*, *49*(3). https://doi.org/10.1509/jmr.09.0315

De-Magistris, T., & Pascucci, S. (2014). The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study. *Economics Letters*, *123*(2). https://doi.org/10.1016/j.econlet.2014.02.016

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*(1). https://doi.org/10.1006/jeem.2001.1199

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2). https://doi.org/10.1257/mic.3.2.114

Essers, B. A. B., van Helvoort-Postulart, D., Prins, M. H., Neumann, M., & Dirksen, C. D. (2010). Does the inclusion of a cost attribute result in different preferences for the surgical treatment of primary basal cell carcinoma?: A comparison of two discrete-choice experiments. *PharmacoEconomics*, *28*(6). https://doi.org/10.2165/11532240-000000000-00000

Feher, M. D., Brazier, J., Schaper, N., Vega-Hernandez, G., Nikolajsen, A., & Bøgelund, M. (2016). Patients' with type 2 diabetes willingness to pay for insulin therapy and clinical outcomes. *BMJ Open Diabetes Research and Care*, *4*(1). https://doi.org/10.1136/bmjdrc-2016-000192

Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? *Transportation Research Part A: Policy and Practice*, *61*. https://doi.org/10.1016/j.tra.2013.12.010

Flores, N. E., & Strong, A. (2007). Cost credibility and the stated preference analysis of public goods. *Resource and Energy Economics*, *29*(3). https://doi.org/10.1016/j.reseneeco.2006.09.001

Gafni, A. (1991). Willingness-to-pay as a measure of benefits: relevant questions in the context of public decision-making about health care programs. *Medical Care*, 29(12), 1246-1252.

Gibson, J. M., Rigby, D., Polya, D. A., & Russell, N. (2016). Discrete Choice Experiments in Developing Countries: Willingness to Pay Versus Willingness to Work. *Environmental and Resource Economics*, *65*(4). https://doi.org/10.1007/s10640-015-9919-8

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Methodological*, *37*(8). https://doi.org/10.1016/S0191-2615(02)00046-2

Grossman, Z., & Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior and Organization*, *84*(2). https://doi.org/10.1016/j.jebo.2012.08.006

Hassan, S., Olsen, S. B., & Thorsen, B. J. (2018). Appropriate Payment Vehicles in Stated Preference Studies in Developing Economies. *Environmental and Resource Economics*, *71*(4). https://doi.org/10.1007/s10640-017-0196-6

Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics (United Kingdom)*, *27*(1). https://doi.org/10.1002/hec.3524

Hensher, D. A. (2007). Attribute Processing in Choice Experiments and Implications on Willingness to Pay. In *Valuing Environmental Amenities Using Stated Choice Studies*. https://doi.org/10.1007/1-4020-5313-4_6

Hensher, D. A., & Greene, W. H. (2010). Non-attendance and dual processing of common-metric attributes in choice analysis: A latent class specification. *Empirical Economics*, *39*(2). https://doi.org/10.1007/s00181-009-0310-x

Hensher, D. A., Rose, J. M., & Greene, W. H. (2012). Inferring attribute non-attendance from stated choice data: Implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation*, *39*(2). https://doi.org/10.1007/s11116-011-9347-8

Hess, S., & Hensher, D. A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological*, *44*(6). https://doi.org/10.1016/j.trb.2009.12.001

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., & Caussade, S. (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). https://doi.org/10.1007/s11116-012-9438-1

Hole, A. R. (2006). Small-sample properties of tests for heteroscedasticity in the conditional logit model. *Economics Bulletin*, *3*(1).

Hynes, S., Hanley, N., & Scarpa, R. (2008). Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American Journal of Agricultural Economics*, *90*(4). https://doi.org/10.1111/j.1467-8276.2008.01148.x

Islam, T., Louviere, J. J., & Burke, P. F. (2007). Modeling the effects of including/excluding attributes in choice experiments on systematic and random components. *International Journal of Research in Marketing*, *24*(4). https://doi.org/10.1016/j.ijresmar.2007.04.002

Jacquemet, N., Joule, R. V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1). https://doi.org/10.1016/j.jeem.2012.05.004

Johnson, F. R., Mohamed, A. F., Özdemir, S., Marshall, D. A., & Phillips, K. A. (2011). How does cost matter in health-care discrete-choice experiments? *Health Economics*, *20*(3). https://doi.org/10.1002/hec.1591

Kardes, F. R., Posavac, S. S., & Cronley, M. L. (2004). Consumer inference: A review of processes, bases, and judgment contexts. *Journal of Consumer Psychology*, *14*(3). https://doi.org/10.1207/s15327663jcp1403_6

Kemper, N., Nayga, R. M., Popp, J., & Bazzani, C. (2016). The Effects of Honesty Oath and Consequentiality in Choice Experiments Nathan. *Agricultural & Applied Economics Association's 2016 AAEA Annual Meeting, Boston, Massachusetts, July 31-August 2*.

Koetse, M. J. (2017). Effects of payment vehicle non-attendance in choice experiments on value estimates and the WTA–WTP disparity. *Journal of Environmental Economics and Policy*, *6*(3). https://doi.org/10.1080/21606544.2016.1268979

Kragt, M. E. (2013). Stated and inferred attribute attendance models: A comparison with environmental choice experiments. *Journal of Agricultural Economics*, *64*(3). https://doi.org/10.1111/1477-9552.12032

Krucien, N., Sicsic, J., & Ryan, M. (2019). For better or worse? Investigating the validity of best–worst discrete choice experiments in health. *Health Economics (United Kingdom)*, *28*(4). https://doi.org/10.1002/hec.3869

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2). https://doi.org/10.1086/259131

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., & Burgess, L. (2013). Best worst discrete choice experiments in health: Methods and an application. *Social Science and Medicine*, *76*(1). https://doi.org/10.1016/j.socscimed.2012.10.007

Louviere, J. J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. A. J. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling*, *1*(1). https://doi.org/10.1016/S1755-5345(13)70025-3

Lusk, J. L. (2003). Effects of cheap talk on consumer willingness-to-pay for golden rice. *American Journal of Agricultural Economics*, *85*(4). https://doi.org/10.1111/1467-8276.00492

Lussier, D. A., & Olshavsky, R. W. (1979). Task Complexity and Contingent Processing in Brand Choice. *Journal of Consumer Research*, *6*(2). https://doi.org/10.1086/208758

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3). https://doi.org/10.1007/BF00133443

Mattmann, M., Logar, I., & Brouwer, R. (2019). Choice certainty, consistency, and monotonicity in discrete choice experiments. *Journal of Environmental Economics and Policy*, *8*(2). https://doi.org/10.1080/21606544.2018.1515118

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*(5). https://doi.org/10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.0.co;2-1

McIntosh, E. (2006). Using discrete choice experiments within a cost-benefit analysis framework: Some considerations. In *PharmacoEconomics* (Vol. 24, Issue 9). https://doi.org/10.2165/00019053-200624090-00004

Mohamed, A. F., Zhang, J., Johnson, F. R., Lomon, I. D., Malvolti, E., Townsend, R., Östgren, C. J., & Parhofer, K. G. (2013). Avoidance of weight gain is important for oral type 2 diabetes treatments in Sweden and Germany: Patient preferences. *Diabetes and Metabolism*, *39*(5). https://doi.org/10.1016/j.diabet.2013.06.001

Orme, B. K. (2010). Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. In *Research Publishers*.

Özdemir, S., Johnson, F. R., & Hauber, A. B. (2009). Hypothetical bias, cheap talk, and stated willingness to pay for health care. *Journal of Health Economics*, *28*(4). https://doi.org/10.1016/j.jhealeco.2009.04.004

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The Adaptive Decision Maker. In *The Adaptive Decision Maker*. https://doi.org/10.1017/cbo9781139173933

Pedersen, L. B., Kjær, T., Kragstrup, J., & Gyrd-Hansen, D. (2011). Does the inclusion of a cost attribute in forced and unforced choices matter? Results from a web survey applying the discrete choice experiment. *Journal of Choice Modelling*, *4*(3). https://doi.org/10.1016/S1755-5345(13)70044-7

Posavac, S. S. (1998). Strategic overbidding in contingent valuation: Stated economic value of public goods varies according to consumers expectations of funding source. *Journal of Economic Psychology*, *19*(2). https://doi.org/10.1016/S0167-4870(98)00004-X

Ratcliffe, J. (2000). The use of conjoint analysis to elicit willingness-to-pay values: Proceed with caution? *International Journal of Technology Assessment in Health Care*, *16*(1). https://doi.org/10.1017/S0266462300161227

Rose, J. M., & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5). https://doi.org/10.1007/s11116-013-9451-z

Ryan, M., & Skåtun, D. (2004). Modelling non-demanders in choice experiments. *Health Economics*, *13*(4). https://doi.org/10.1002/hec.821

Ryan, M., Watson, V., & Entwistle, V. (2009). Rationalising the "irrational": A think aloud study of discrete choice experiment responses. *Health Economics*, *18*(3). https://doi.org/10.1002/hec.1369

Sælensminde, K. (2006). Causes and consequences of lexicographic choices in stated choice studies. *Ecological Economics*, *59*(3). https://doi.org/10.1016/j.ecolecon.2005.11.001

Scarpa, R., & Thiene, M. (2005). Destination choice models for rock climbing in the North eastern Alps: A latent-class approach based on intensity of preferences. *Land Economics*, *81*(3). https://doi.org/10.3368/le.81.3.426

Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1). https://doi.org/10.1093/ajae/aas073

Schotter, A. (2003). Decision making with naive advice. *American Economic Review*, *93*(2). https://doi.org/10.1257/000282803321947047

Sever, I., Verbič, M., & Klarić Sever, E. (2019). Cost attribute in health care DCEs: Just adding another attribute or a trigger of change in the stated preferences? *Journal of Choice Modelling*, *32*. https://doi.org/10.1016/j.jocm.2018.03.005

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*, *41*(22). https://doi.org/10.1080/00036840801964633

Silva, A., Nayga, R. M., Campbell, B. L., & Park, J. L. (2012). Can perceived task complexity influence cheap talk's effectiveness in reducing hypothetical bias in stated choice studies? *Applied Economics Letters*, *19*(17). https://doi.org/10.1080/13504851.2012.667532

Smith, R. D. (2003). Construction of the contingent valuation market in health care: A critical assessment. In *Health Economics* (Vol. 12, Issue 8). https://doi.org/10.1002/hec.755

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete Choice Experiments in Health Economics: Past, Present and Future. In *PharmacoEconomics* (Vol. 37, Issue 2). https://doi.org/10.1007/s40273-018-0734-2

Stevens, T. H., Tabatabaei, M., & Lass, D. (2013). Oaths and hypothetical bias. *Journal of Environmental Management*, *127*. https://doi.org/10.1016/j.jenvman.2013.04.038

Swait, J., & Adamowicz, W. (2001). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organisational Behavior and Human Decision Processes*, *86*(2). https://doi.org/10.1006/obhd.2000.2941

Train, K. E. (2009). Discrete choice methods with simulation, second edition. In *Discrete Choice Methods with Simulation, Second Edition* (Vol. 9780521766555). https://doi.org/10.1017/CBO9780511805271

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., & de Wit, G. A. (2014). The effect of including an opt-out option in discrete choice experiments. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0111805

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4). https://doi.org/10.1257/mic.4.4.145

Zawojska, E., Bartczak, A., & Czajkowski, M. (2019). Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management*, *93*. https://doi.org/10.1016/j.jeem.2018.11.007

Zawojska, E., & Czajkowski, M. (2017). Re-examining empirical evidence on stated preferences: importance of incentive compatibility. *Journal of Environmental Economics and Policy*, *6*(4). https://doi.org/10.1080/21606544.2017.1322537

Figure captions

Figure 1. Illustration of a choice task (COST DCE version*)

Figure 2. Comparison of relative importance scores – COST versus NOCOST DCEs

Please compare the three support services (A, B, and C) and then answer the question below by clicking on the button for the service you choose. Each support service would be provided once a week for six weeks.

Service A:
- Provides everyone with the same information
- Makes suggestions that fit your current situation
- Seems to think that everyone wants to get the same from life
- Communicates in a friendly and personal way
- Costs £10 per week

Service B:
- Provides information that is relevant to you
- Makes suggestions that fit your current situation
- Seems to think that everyone wants to get the same from life
- Communicates in a neutral professional way
- Costs £10 per week

Service C:
- Provides information that is relevant to you
- Takes little account of your current situation
- Works with you on what you want to get from life
- Communicates in a friendly and personal way
- Costs £20 per week

1. Which service would you like the **most?**

   Service A ☐ Service B ☐ Service C ☐

2. Would you actually **buy** your **most** preferred option?

   YES ☐ NO ☐

*The NOCOST condition was identical in all respects other than the COST attribute was excluded.

Figures (NO AUTHOR DETAILS)

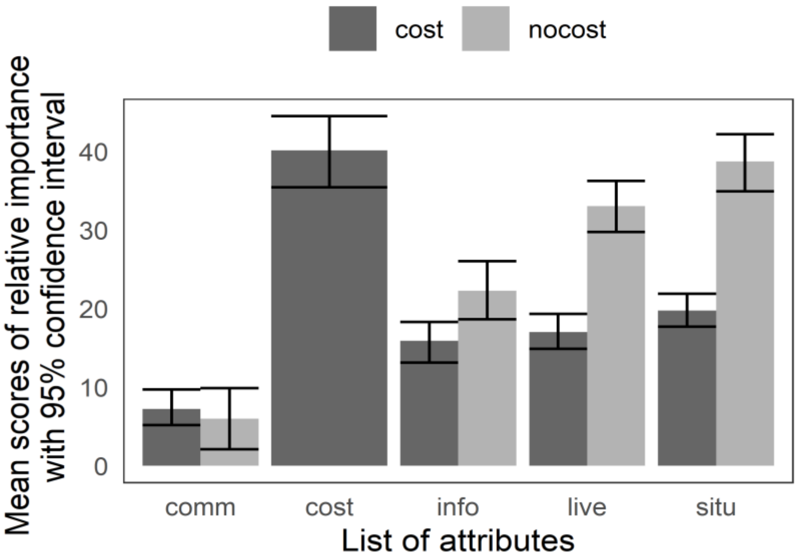

Table 1. Attributes and levels used to describe personalisation of chronic pain self-management services

Attributes (Label)	Description	Levels
Information (INFO)	Information about pain, the conditions that cause it, and the different ways there are of managing it	Provides everyone with the same information Provides information that is relevant to you*
Situation (SITU)	Things like where you live, who you live with, what resources you have, what you usually do for yourself and others, and how pain currently affects that	Takes little account of your current situation Makes suggestions that fit your current situation*
Living well (LIVE)	Things that really matter to you, especially the kinds of things that you would like to achieve or to spend more time doing, and the kind of person that you want to be	Seems to think that everyone wants to get the same from life Works with you on what you want to get from life*
Communication (COMM)	The way that the support service might communicate with you	Communicates with you in a neutral professional way Communicates with you in a friendly and personal way*
Cost (COST)**		£5, £10, £15, £20

* Level corresponding to a higher level of personalisation

** Included only in the COST condition. The following instructions were given: "*We are interested in how you would value the different support services. One way of doing this is to ask about the amount of money you would be willing to pay for them. In the choice questions that follow, each support service has a cost. Please assume that the support services are not available on the NHS so you would have to pay this amount. We understand that some of the choices may be difficult to make, but there are no right or wrong answers. Your personal opinion is what matters*".

Table 2. Respondents' characteristics and experience across the COST and NOCOST experiments

	COST DCE (N=517)	NO COST DCE (N=206)	p-value
Marital status			0.285
Not single	64.4	68.9	
Single	35.6	31.1	
EDUCATION			0.753
Other than university	57.4	55.8	
College/university	42.6	44.2	
GENDER			0.001
Male	34	48.1	
Female	66	51.9	
AGE			0.516
AGE1 (< 50 years)	31.1	30.1	
AGE2 (50-60 years)	26.9	31.1	
AGE3 (> 60 years)	42	38.8	
Annual INCOME			0.201
INCOME1 (< £ 15,600)	28.2	22.3	
INCOME2 (£ 15,000-£ 31,200)	34	33.5	
INCOME3 (> £ 31,200)	27.3	34.5	
INCOME4 (No say)	10.4	9.7	
EXPERIENCE of a paying CPSM service during the past six months? (CREDIBILITY)			0.628
No	62.5	64.6	
Yes	37.5	35.4	
When you were making your choices, did you have any particular type of support service in mind? (PERCEIVED REALISM)			0.235
No	52.8	47.6	
Yes	47.2	52.4	

CPMS: Chronic pain management services; p-values based on Chi-square test comparing proportions between COST and NOCOST samples.

Table 3. Stated attribute non-attendance - COST and NOCOST experiments

Attributes				**p-value**
INFORMATION	*I always considered*	*I sometimes considered*	*I never considered*	
COST	67.50%	29.21%	3.29%	
NOCOST	67.48%	31.08%	1.46%	0.376
SITUATION				
COST	66.92%	30.17%	2.90%	
NOCOST	72.33%	24.27%	3.40%	0.28
LIVING WELL				
COST	57.06%	39.46%	3.48%	
NOCOST	62.62%	33.49%	3.88%	0.328
COMMUNICATION				
COST	47%	46.62%	6.38%	
NOCOST	53.40%	41.75%	4.85%	0.275
COST				
COST	64.60%	25.33%	10.06%	NA
NOCOST	-	-	-	-

Chi-square tests were used to test differences between COST and NOCOST conditions, NA=not applicable

Table 4. MNL, heteroscedastic MNL, and mixed logit models testing for the effect of the cost attribute on the error variance (consistency)

	MNL COST DCE	MNL NOCOST DCE	HMNL (COST +NOCOST)	HMXL (COST + NOCOST)#	
Var.	Coefficient (St.err.)	Coefficient (St.err.)	Coefficient (St.err.)	Coefficient (St.err.)	SD (St.err.)
ASC-B	0.283 (0.061) ***	0.305 (0.075) ***	0.321 (0.053) ***	0.412 (0.062) ***	-
ASC-C	0.010 (0.063)	0.065 (0.077)	0.029 (0.055)	0.049 (0.060)	-
ASC-NOBUY	2.415 (0.090) ***	3.126 (0.117) ***	3.046 (0.118) ***	3.256 (0.250) ***	3.950 (0.281) ***
INFORMATION	0.702 (0.050) ***	0.759 (0.063) ***	0.816 (0.050) ***	1.041 (0.087) ***	0.962 (0.091) ***
SITUATION	0.871 (0.055) ***	1.320 (0.075) ***	1.198 (0.065) ***	1.765 (0.126) ***	1.259 (0.114) ***
LIVING WELL	0.751 (0.054) ***	1.126 (0.072) ***	1.032 (0.061) ***	1.422 (0.103) ***	0.873 (0.099) ***
COMMUNICATION	0.318 (0.049) ***	0.203 (0.062) ***	0.305 (0.044) ***	0.331 (0.065) ***	0.744 (0.087) ***
COST##	-0.088 (0.005) ***	-	-0.116 (0.009) ***	-1.284 (0.104) ***	1.856 (0.135) ***
Covariates of scale					
COSTDCE (Cost effect)	-	-	-0.304 (0.049) ***	-0.226 (0.063) ***	
Gender (Female)	-	-	0.076 (0.029) ***	0.048 (0.060)	
Model diagnostics					
LL at convergence	-5558.974	-2764.1	-8329.157	-5974.945	
McFadden's pseudo- R^2	0.068	0.101	0.093	0.349	
AIC/n	1.795	2.242	1.922	1.381	
BIC/n	1.803	2.258	1.93	1.394	
n (observations)	6204	2472	8676	8676	
r (respondents)	517	206	723	723	
k (parameters)	8	7	10	16	

**** p<0.01, ** p<0.05. Standard errors in bracket. ASC=Alternative Specific Constant*

#We refer to the heteroskedastic MXL (HMXL) model as an MXL model in which scale is allowed to systematically differ between the COST and NOCOST DCE samples. This is a natural extension of an HMNL model.

##In the HMXL model, the cost coefficient was specified to be log-normally distributed. Abbreviation: St.err. , Standard error

Table 5. Respondents' preferences classified into latent classes - the results of the unconstrained latent class logit model

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
	Cost ignorers	**Cost ignorers**	**Cost compensators**	**Cost compensators**	**Cost minimisers**	**Cost compensators**
Preference parameters						
ASC-B	1.086 (0.302) ***	1.143 (1.188)	0.173 (0.191)	0.173 (0.112)	0.192 (0.237)	0.397 (0.419)
ASC-C	0.488 (0.374)	0.834 (1.819)	-0.045 (0.195)	-0.114 (0.106)	-0.210 (0.272)	0.269 (0.345)
ASC-NOBUY	4.068 (0.536) ***	8.036 (2.097) ***	2.517 (0.295) ***	-0.760 (0.320) **	-2.194 (0.505) ***	0.789 (0.439) *
INFORMATION	0.838 (0.229) ***	0.916 (1.108)	1.133 (0.136) ***	0.924 (0.067) ***	-0.073 (0.292)	0.926 (0.307) ***
SITUATION	1.077 (0.227) ***	3.559 (1.742) ***	2.040 (0.172) ***	1.174 (0.110) ***	-0.263 (0.258)	1.880 (0.322) ***
LIVING WELL	0.798 (0.233) ***	1.540 (1.226)	1.835 (0.175) ***	1.090 (0.102) ***	-0.179 (0.184)	0.695 (0.384) *
COMMUNICATION	0.233 (0.215)	0.065 (0.763)	0.678(0.147) ***	0.484 (0.088) ***	-0.039 (0.156)	0.745 (0.316) **
COST	**-0.029 (0.026)**	**-0.300 (0.155)**	**-0.135 (0.016) *****	**-0.036 (0.007) *****	**-0.177 (0.024) *****	**-0.466 (0.057) *****
Membership parameters						
CONSTANT	-0.511 (0.375)	1.332 (0.198) ***	0.183 (0.250)	0.026 (0.236)	-1.067 (0.355) ***	0.000 (fixed)
RESPONSE TIME	-0.308 (1.091)	-0.476 (0.248)	-0.098 (0.342)	-0.435 (1.050)	-3.320 (1.205) ***	0.000 (fixed)
PAY EXPERIENCE (vs. NO)	**-0.770 (0.326) ****	**-0.459 (0.162) *****	**-0.170 (0.208)**	**-0.160 (0.193)**	**-0.537 (0.347)**	0.000 (fixed)
SERVICE IN MIND (vs. NO)	0.017 (0.260)	-0.444 (0.178) **	-0.142 (0.227)	0.082 (0.217)	0.363 (0.332)	0.000 (fixed)
INCOME2 (£ 15,600 - £ 31, 199)	0.336 (0.338)	0.067 (0.193)	0.245 (0.242)	0.169 (0.257)	-0.029 (0.366)	0.000 (fixed)
INCOME3 (> £31, 199)	0.895 (0.332) ***	0.032 (0.205)	0.263 (0.259)	0.527 (0.235) **	0.272 (0.314)	0.000 (fixed)
CLASS SHARE (%)	8.553	46.512	13.748	12.57	6.022	12.595
Model diagnostics						
LL at convergence	-3737.532					
McFadden's pseudo-R2	0.373					
BIC/n	1.315					
n (observations)	6204					
r (respondents)	517					
k (parameters)	78					

***, *** indicate significance at 5%, and 1% level, respectively. Standard errors given in parentheses. INCOME is also included as an additional control variable*