ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Mimicking Real-Life Decision Making in Health: Allowing Respondents Time to Think in a Discrete Choice Experiment

Jorien Veldwijk, PhD,[1,2,*] Jennifer Viberg Johansson, PhD,[2] Bas Donkers, PhD,[3] Esther W. de Bekker-Grob, PhD[1]

[1]*Erasmus School of Health Policy & Management, and Erasmus Choice Modelling Centre, Erasmus University, Rotterdam, The Netherlands;* [2]*Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden;* [3]*Erasmus School of Economics, and Erasmus Choice Modelling Centre, Erasmus University, Rotterdam, The Netherlands.*

ABSTRACT

Objective: To empirically test the impact of allowing respondents time to think (TTT) about their choice options on the outcomes of a discrete choice experiments (DCE).

Methods: In total, 613 participants of the Swedish CArdioPulmonary bioImage Study (SCAPIS) completed a DCE questionnaire that measured their preferences for receiving secondary findings of a genetic test. A Bayesian D-efficient design with 60 choice tasks divided over 4 questionnaires was used. Each choice task contained 2 scenarios with 4 attributes: type of disease, disease penetrance probability, preventive opportunities, and effectiveness of prevention. Respondents were randomly allocated to the TTT or no TTT (NTTT) sample. Latent class models (LCMs) were estimated to determine attribute-level values and their relative importance. In addition, choice certainty, attribute-level interpretation, choice consistency, and potential uptake rates were compared between samples.

Results: In the TTT sample, 92% of the respondents (245 of 267) indicated they used the TTT period to (1) read the information they received (72%) and (2) discuss with their family (24%). In both samples, respondents were very certain about their choices. A 3-class LCM was fitted for both samples. Preference reversals were found for 3 of the 4 attributes in one class in the NTTT sample (34% class-membership probability). Relative importance scores of the attributes differed between the 2 samples, and significant scale effects indicating higher choice consistency in TTT sample were found.

Conclusions: Offering respondents TTT influences decision making and preferences. Developers of future DCEs regarding complex health-related decisions are advised to consider this approach to enhance the validity of the elicited preferences.

Keywords: discrete choice experiment, hypothetical bias, stated preferences, time-to-think.

VALUE HEALTH. 2020; ■(■):■–■

Introduction

To determine individuals' or patients' preferences for healthcare treatments and services, a discrete choice experiment (DCE) can be conducted.[1–3] DCEs are increasingly being used in healthcare to inform policy makers about individuals' or patients' preferences and the associated welfare measures such as willingness-to-pay or willingness-to-accept risks that can be calculated based on those preferences.[4–7] Because DCEs can be used as input for policy making, the accuracy and validity of the measured (ie, stated) preferences are essential.

There are 2 issues to consider, given that respondents of DCEs are asked to choose immediately between hypothetical situations within a highly controlled situation without any constraints (ie, respondents do not actually have to spend their money or time based on their choices).[8] First, in the absence of any constraints, respondents' stated preferences are likely to differ from their revealed preferences because they are valuing hypothetical situations and neglect or undervalue the importance of specific program characteristics (hypothetical bias).[2,9–12] For instance, respondents often overestimate their willingness to pay for a treatment,[12–14] whereas they underestimate their willingness to accept risks.[15] Second, health- and treatment-related decisions in real life are largely influenced by psychosocial factors such as emotions, perceived susceptibility, and social norms,[16–18] whereas DCEs generally do not accommodate for these effects as they solely focus on the valuation of treatment attributes.

To potentially reduce the impact of hypothetical bias and the influence of psychosocial factors at once, a time-to-think (TTT) approach may be used.[10,19,20] In a TTT-DCE, respondents are asked

* Address correspondence to: Jorien Veldwijk, Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Email: veldwijk@eshpm.eur.nl

https://doi.org/10.1016/j.jval.2020.02.014

to read all of the information in the questionnaire along with the choice tasks, without completing it. They can complete the questionnaire at least 1 day later. This gives them the opportunity to think about the information they received; to deliberate about choice options; to discuss the topic with their partner, family, or friends; or to find/read additional information. The TTT approach mimics real-life decision making in healthcare to a greater extent, which is especially valuable given the complexity of treatment decisions in healthcare DCEs.[21] It is expected that allowing respondents TTT will lead to more careful deliberation of the attributes and discussion with their peers, thereby reducing the influence of hypothetical bias and psychosocial factors. Therefore, a TTT approach potentially adds to the validity of DCE outcomes.

A TTT approach has been shown to be effective in environmental economics and interview-led DCE studies.[10,15,22,23] Nevertheless, evidence is lacking on whether this is a workable and valuable approach in current state-of-the-art online DCE studies conducted in healthcare. Therefore, this study aims to empirically test whether and to what extent the outcomes of a DCE change when respondents are allowed TTT about their choice options.

Methods

DCE Case Study and Participant Recruitment

A DCE measuring research participants' preferences to receive secondary findings of a genetic test was used as a case for this study.[24] Although such decisions are considered highly complicated,[21,25,26] this makes it an interesting case for this study. A total of 1300 participants of the Swedish CArdioPulmonary bioImage Study (SCAPIS)[27] were selected to receive the DCE questionnaire. Invitations to participate were sent via postal mail and contained a personalized link to the online survey that could be completed at home. The SCAPIS research program is a cohort study including 30 000 Swedes aged 50 to 64 years, which aims to find risk factors to predict cardiopulmonary disease risk levels among its participants.[27] The Regional Ethical Review Board in Gothenburg approved the study (Dnr: 610-16) in accordance with the guidelines laid down in the Declaration of Helsinki.

Attributes, Levels, and Choice Task Presentation

Attributes and levels were identified based on previously published literature,[28–35] 4 focus group interviews with a total of 16 SCAPIS research participants, and in-depth discussion with 5 experts (eg, clinical geneticists, genetic counselor). Finally, 4 attributes were selected for this DCE (Table 1). A more detailed description of the attribute selection process is described in Viberg Johansson et al.[24]

Experimental Design and Pilot Testing

NGene 1.0 (ChoiceMetrics, 2011) software was used to construct a Bayesian D-efficient design.[9,36] The design of the pilot DCE was constructed using priors indicating the direction (ie, positive or negative) of the included attribute-level estimates. The draft questionnaire was pilot tested among a subgroup (n = 22) of our study population to check whether correct wording was used and whether the target population understood the choice tasks. Eight of these pilot tests were "think-aloud" tests, in which a researcher was present while the participant completed the questionnaire by reading aloud and expressing their thoughts. Based on the pilot tests, minor changes were made throughout the questionnaire related to layout and wording.

The final DCE design was created based on a multinomial-logit model, including priors from the previously described pilot study. Based on outcomes of the interviews with experts, 2-way interactions between "type of the disease" with "disease penetrance probability" and "effectiveness of the preventive measure" were included. The final design consisted of 60 unique choice tasks, divided over 4 blocks. Respondents were randomly assigned to 1 of these blocks.

Questionnaire

The first section of the questionnaire contained questions on demographics and respondents' general health status. The second part of the questionnaire contained the choice tasks of the DCE. Respondents received training on the attributes and levels (see Appendix 1 in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.02.014) illustrated by 2 examples. Each choice task started with the question, "Imagine that you can get additional genetic risk information from participating in a SCAPIS genetic test. In which situation would you prefer to receive such information, situation 1 or situation 2?" After every choice task, participants were asked to indicate how certain they were about the choice they made on a 7-point Likert-type scale (ranging from *very certain* to *very uncertain*). In the third part, health literacy[37]

Table 1. Attributes and levels of direct choice experiment.

Attribute	Level 1	Level 2	Level 3	Level 4	Level 5
Type of the disease: Different types of diseases might have a different impact on you. Genetic tests can be used to identify different types of diseases and conditions.					
	Life threatening	Physical disease	Mental disease	Physical disability	
Disease penetrance: If the results from the genetic test show that you are at risk of a disease, it is not certain that you will get the disease. The probability of getting the disease can be explained by how many people out of every 100, that have the same type of gene variant as you, will be affected.					
	5 out of every 100	30 out of every 100	80 out of every 100		
Preventive opportunities: Some diseases are preventable. By taking action, you can decrease the risk of getting the disease. It does not mean that the risk disappears, but it can decrease.					
	Nothing	Operation	Medication	Lifestyle changes	
Effectiveness of the preventive measure: Different preventive measures have different abilities to reduce the risk of getting the disease. The number of individuals who will not get the disease because of prevention is calculated based on how many got the disease in the first place. For example, if 30 persons out of every 100 will get the disease and all people take the preventive measure, with an effectiveness of 50%, 15 of 30 will not get the disease anymore.					
	0%	25%	50%	75%	90%

and numeracy[38] were measured. In addition, respondents who completed the TTT-DCE were asked if and how they had used their TTT period. The complete survey was web based and constructed in Sawtooth Software SSI Web 8.4.8.

Respondents were randomly assigned to either the TTT or no TTT (NTTT) questionnaire. Respondents assigned to the NTTT sample were asked to answer all questions of the questionnaire at once, as is common practice in survey research. Respondents assigned to the TTT arm were asked to complete the first part of the questionnaire and to read all the information (including the choice task examples) regarding the DCE. After that, respondents were informed that their survey ended as well as that they would be given time to consider the decisions that they were going to make. Respondents could reenter the survey at any time, and they were informed that they would receive a request to complete the rest of the questionnaire after 7 days. When entering the survey after these 7 days, respondents were offered the opportunity to read all DCE-related material once again. All respondents who did not start the questionnaire as well as respondents who started the TTT part 1 but did not respond to the request to complete the rest of the survey received 2 reminders with a 2-week interval.

Statistical Analysis

Sample descriptions

The TTT and NTTT samples are described based on mean age, gender, educational level, health literacy and numeracy mean score, and self-perceived general health status. By means of independent-sample *t* tests and chi-square tests, differences in these variables between the TTT and NTTT sample, between TTT completers and TTT dropouts, and between respondents of all 4 blocks in both the TTT and NTTT sample were tested.

Dominant decision-making behavior

The proportion of respondents who always chose the alternative with the highest penetration probability, the highest level of effectiveness of the preventive measure, or the opt-out was calculated for each data set. Chi-square tests were conducted to test whether this differed significantly between the TTT and NTTT samples.

In addition, the mean number of times respondents in both samples chose the alternative with the highest penetration probability, the highest effectiveness of the preventive measure, or the opt-out was calculated. Independent-sample *t* tests were used to test whether the means differed significantly between the TTT and NTTT samples.

Choice certainty

A mean certainty score was calculated for all participants. Differences in the mean certainty score between the TTT and NTTT sample were calculated via an independent-sample *t* test. In addition, the proportion of respondents who indicated to be (very) certain about their choices was calculated in both samples, and differences in these proportions were estimated by means of chi-square tests.

Attribute-level interpretation

Part-worth utilities for the 2 numerical attributes (ie, disease penetrance probability and effectiveness of the preventive measure) were estimated to analyze whether respondents interpreted them correctly on a linear scale. For that purpose, a dummy variable representing the highest level of the attributes was added to the linear attribute parameter in the regression model. The difference in part-worth utilities between level 1 and 2 is significantly different from the part-worth utilities between level 2 and 3 (indicating a nonlinear attribute), if the dummy variable is significant.

Attribute estimates and relative importance

Panel latent class models (LCMs) were estimated to adjust for the multilevel structure of the data and to detect preference heterogeneity.[39] In this model, all attributes, except "effectiveness of the preventive measure," were determined to be nonlinear and therefore recoded using effects codes.[9,40] Based on model fit tests (AIC, log-likelihood), the model most suitable for our data was selected (see equation below), and how many classes could be identified within the data (models ranging from 1 to 4 classes) was tested.

$$V_{rta|c} = \beta0_{|c} + \beta1_{|c}\ \text{type of disease}_{\text{physical disease } rta|c} + \beta2_{|c}\ \text{type of disease}_{\text{mental disease } rta|c} + \beta3_{|c}\ \text{type of disease}_{\text{physical disability } rta|c} + \beta4_{|c}\ \text{disease penetrance probability}_{\text{30 out of 100 } rta|c} + \beta5_{|c}\ \text{disease penetrance probability}_{\text{80 of 100 } rta|c} + \beta6_{|c}\ \text{preventive opportunity}_{\text{operation } rta|c} + \beta7_{|c}\ \text{preventive opportunity}_{\text{medication } rta|c} + \beta8_{|c}\ \text{preventive opportunity}_{\text{lifestyle changes } rta|c} + \beta9_{|c}\ \text{effectiveness of preventive measure}_{rta|c}$$

$$V_{opt-out} = 0$$

The systematic utility component (V) describes the observable utility that respondent ® belonging to class (c) reported for alternative (a)a in choice task (t)t. β_0 represents the alternative specific constant for receiving secondary findings from a genetic test with 0% effectiveness and the average utility level for all other attributes, and β_1 to β_9 are the attribute-level estimates. Interaction terms between the attributes were not included, as they were identified to be insignificant, confirming the outcomes of the focus groups with the target population.

Importance weights for the attributes relative to the most important attribute were calculated based on the results of the LCMs, separately for all classes. The class-adjusted relative importance was calculated by computing the relative importance score of all attributes in each class separately as described above, after which they were weighted according to class assignment probability.

Choice consistency

To test whether differences between the TTT and NTTT were attributable to any significant differences in the scale parameter (ie, if choice consistency between respondents is significantly higher in one sample compared with the other), a heteroscedastic MNL (HMNL) model was estimated on the pooled data set.[1,41]

Participation probabilities

Separately for the TTT and NTTT samples, the probability that individuals are willing to receive secondary finding from a genetic test was estimated for different realistic implementation scenarios. Interclass correlations were estimated to measure the level of agreement between the probabilities of every estimated implementation scenario.

Results

Sample Description

Within the NTTT sample, 391 of the 650 questionnaires (60%) were returned. After removing respondents with >10% missing answers on their choice task, 346 of these questionnaires (88%) were used in the analysis. Within the TTT sample, initially, 381 of the 650 questionnaires (59%) were returned. Of these, 287 (75%) also returned their questionnaire after the TTT period, indicating 94 respondents (25%) dropped out. After removing respondents with >10% missing answers on their choice task, 267 of these questionnaires (93%) were used in the analysis.

Demographic variables did not differ between the NTTT and TTT sample (Table 2), nor did they differ between the respondents who completed the TTT-DCE and those who dropped out or between respondents who completed the different blocks of the DCE.

Of the respondents who completed the TTT-DCE, 92% indicated they had used the TTT period to think about the choices they were offered and the information they had received (Table 2). Most (72%) indicated they read the information they received again, 24% of the respondents reported that they discussed the matter with their family, and 15% indicated they did something else (which in most cases was described as self-reflection/deliberation).

Choice Certainty

Within both samples, respondents reported a high mean certainty score (5.5 of 7 [SD 1.0]), which did not (significantly) differ between samples. The proportion of respondents who indicated they were very certain about their choice was also equal in both samples (34%).

Dominant Decision-Making Behavior

In the TTT and NTTT sample, a relatively small proportion of the respondents showed dominant decision-making behavior (ie, lexicographic preferences) for choosing the opt-out (2% and 1%, respectively) and for choosing the alternative with the highest disease penetrance probability (1% and 2%, respectively). No lexicographic preferences were found for the effectiveness of the preventive measure attribute. The mean number of times the opt-out was chosen (TTT 2.1 [SD 3.6]; NTTT 1.6 [SD 3.0]) and the mean number of times the alternative with the highest disease penetrance probability was chosen (TTT 5.8 [SD 2.7]; NTTT 5.2 [SD 2.7]) were significantly higher in the TTT sample as compared with the NTTT sample.

Table 2. Demographics and TTT-period activities stratified by sample

	TTT	NTTT
Mean (SD) age, y	58.8 (4.5)	58.8 (4.3)
Gender		
Male	39%	43%
Female	61%	57%
Educational level		
Primary school	8%	11%
High school	31%	30%
University or higher education	61%	59%
Mean health literacy score (SD)*	3.8 (0.6)	3.8 (0.7)
Mean health numeracy score (SD)	4.7 (0.91)	4.6 (1.02)
General health status		
Poor	2%	2%
Average	21%	15%
Good	77%	84%
Proportion of respondents who reported they used the TTT period	92%	N/A
Activities undertaken		
Read the information over	72%	N/A
Searched on internet	6%	N/A
Discussed with family	24%	N/A
Discussed with friends	9%	N/A
Other	15%	N/A

TTT indicates time to think.
*Problematic health literacy for mean scores between 1.8 and 2.4, inadequate health literacy for mean scores between 0 and 1.6 points.[36,37]

Attribute-Level Interpretation

In both the TTT and NTTT sample, the dummy variable indicating the highest level for disease penetrance probability was shown to be statistically significant (TTT $\beta = -.55$; $P = < .01$, NTTT $\beta = -.45$; $P < .01$), indicating respondents valued the part-worth utilities between the levels nonlinearly. Given the levels of this attribute, this shows that even though the distance between 30 and 80 is about twice as large as the difference between 5 and 30, the impact of moving from 30 to 80 on utility is smaller compared with moving from 5 to 30. The effectiveness of the preventive measure attribute was considered to be valued as a continuous attribute and therefore included as a linear variable in further analyses (TTT $\beta = -.18$; $P = .03$, NTTT $\beta = .05$; $P = .48$).

Attribute-Level Estimates and Relative Importance

In both the TTT and NTTT samples, a 3-class LCM turned out to be the best-fitting model, indicating significant preference heterogeneity (Table 3). Classes are numbered according to class size, with class 1 being the largest.

In both samples, the alternative specific constant in class 3 was negative, indicating an a priori disutility for receiving secondary findings from a genetic test with 0% effectiveness and the average utility level for all other attributes, whereas the alternative specific constant for the other 2 classes was positive (Table 3). In both samples, the positive attribute-level estimate for effectiveness of the preventive intervention indicated respondents prefer to receive secondary findings from a genetic test if the effectiveness of the available preventive measure increases.

A preference reversal was found when comparing class 2 of the NTTT sample to the other classes in the NTTT as well as the TTT sample. The most pronounced differences were found for the disease penetrance probability attribute and the preventive opportunity attribute. Respondents in NTTT class 2 preferred to receive secondary findings from a genetic test if the disease penetrance probability decreases (as compared with increases) and when there are no preventive opportunities available (as compared with any of the availability preventive measures). The probability of belonging to this class was 34%.

The effectiveness of the preventive measures attribute was most important in both samples; however, the class-adjusted relative importance scores differed between the TTT and NTTT samples (Fig. 1).

Choice Consistency

The HMNL model showed a positive significant scale factor when comparing the TTT and NTTT data files ($P = .00$). This indicates that, as compared with the NTTT sample, respondents in the TTT sample were more consistent when answering their choice tasks.

Participation Probabilities

The participation probabilities were calculated for some realistic scenarios of returning secondary findings of a genetic test (Table 4). Although the probabilities were relatively similar, the mean utility and participation probability calculated based on the TTT data were always lower as compared with those based on the NTTT data. The interclass correlation coefficient—when comparing the average class-adjusted participation probabilities between both samples—was 0.97.

Table 3. Preferences for receiving secondary findings based on a panel latent class analysis separately for the TTT and NTTT sample.

	TTT sample						NTTT sample					
	Class 1		Class 2		Class 3		Class 1		Class 2		Class 3	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Constant	1.11‡	0.27	1.88‡	0.28	−2.93‡	0.27	2.98‡	0.25	1.64‡	0.38	−1.97‡	0.16
Type of disease												
Life threatening (ref)	0.47‡	0.10	1.24‡	0.12	0.03	0.14	1.33‡	0.09	0.28†	0.13	0.26†	0.11
Physical disease	−0.09	0.10	-0.35‡	0.11	−0.09	0.15	−0.39‡	0.08	0.03	0.11	0.16	0.11
Mental disease	−0.65‡	0.12	−0.12	0.10	0.06	0.14	−0.39‡	0.08	−0.88‡	0.15	−0.23†	0.11
Physical disability	0.27†	0.10	−0.77‡	0.14	−0.00	0.14	−0.55‡	0.10	0.57‡	0.12	−0.19*	0.10
Disease penetrance probability												
5 out of 100 (ref)	−0.54‡	0.10	−1.37‡	0.10	−0.61‡	0.12	−1.21‡	0.09	0.55‡	0.13	−0.59‡	0.09
30 out of 100	0.22‡	0.07	0.08	0.08	−0.03	0.12	0.17‡	0.06	0.15*	0.08	−0.16*	0.08
80 out of 100	0.32‡	0.12	1.29‡	0.12	0.64‡	0.12	1.04‡	0.12	−0.70‡	0.18	0.75‡	0.09
Preventive opportunities												
None (ref)	−0.33‡	0.11	−1.69‡	0.12	−0.14	0.18	−1.87‡	0.10	0.38‡	0.14	−0.74‡	0.11
Operation	−0.07	0.12	0.62‡	0.17	−0.62‡	0.23	0.57‡	0.14	−0.36†	0.15	−0.13	0.13
Medication	−0.17*	0.10	0.30‡	0.10	0.02	0.17	0.58‡	0.09	−0.36‡	0.14	0.11	0.11
Lifestyle changes	0.57‡	0.11	0.77‡	0.10	0.74‡	0.15	0.72‡	0.07	0.34†	0.14	0.76‡	0.10
Effectiveness of the preventive measure	0.39‡	0.04	0.18‡	0.03	0.28‡	0.04	0.20‡	0.03	0.48‡	0.08	0.25‡	0.03
Mean class probability	0.43		0.38		0.19		0.44		0.34		0.22	

Note. The attribute-level estimate of the reference categories can be calculated as −1 × sum of the other attribute level estimates.
SE indicates standard error.
*$P < .10$.
†$P < .05$.
‡$P < .01$.

Conclusions and Discussion

This study aimed to empirically test whether and to what extent the outcomes of a DCE differ when respondents are allowed TTT about their choice options. Results showed that, irrespective of whether respondents were offered TTT, they indicated they were very certain about the choices that they made in the DCE. Nevertheless, the significant scale factor in the HMNL model showed that respondents in the NTTT sample were significantly less consistent in their choices as compared with respondents in the TTT sample, which is in line with previous findings.[22] In addition, respondents in the NTTT sample undervalued the attribute related to disease penetrance probability, which is the most abstract and complicated attribute in this DCE.

Figure 1. Relative importance scores adjusted for class assignment separately for the time to think (TTT) and no time to think (NTTT) samples.

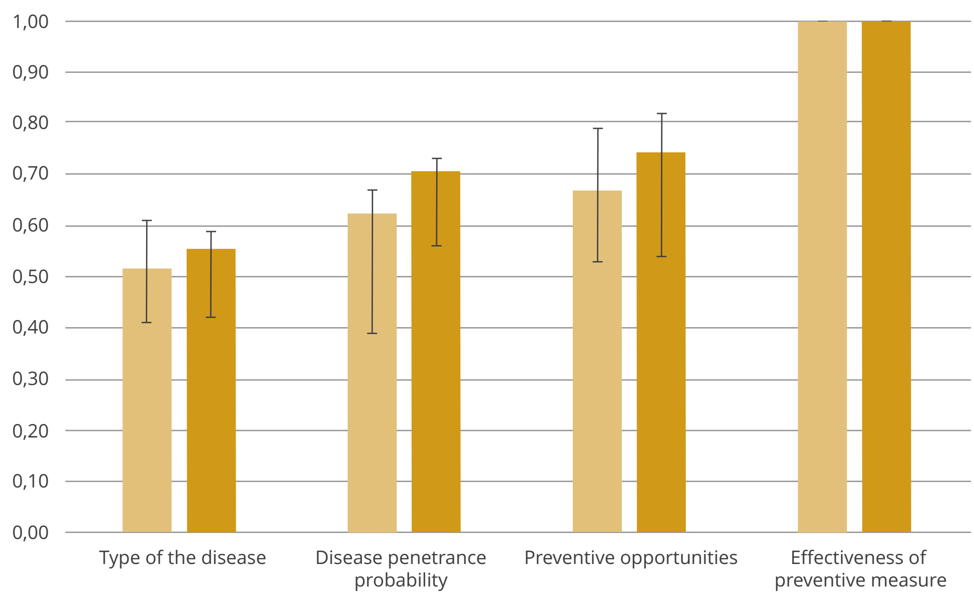

Table 4. Utility and probability scores for receiving secondary findings given realistic implementation scenarios separate for the time to think (TTT) and no time to think (NTTT) samples.

	TTT				NTTT			
	Class 1	Class 2	Class 3	Average	Class 1	Class 2	Class 3	Average
Life-threatening disease for which an operation is possible								
Scenario 1: • Penetrance: 5 out of 100 • Effectiveness: 25%								
Utility	1.95	2.82	−3.43	1.26	4.17	3.31	−1.81	2.56
Probability	0.87	0.94	0.03	0.74	0.98	0.96	0.14	0.79
Scenario 2: • Penetrance: 80 out of 100 • Effectiveness: 25%								
Utility	2.81	5.48	−2.18	2.87	6.42	2.06	−0.47	3.42
Probability	0.94	1.00	0.10	0.80	1.00	0.89	0.39	0.83
Scenario 3: • Penetrance: 5 out of 100 • Effectiveness: 90%								
Utility	4.22	1.68	−1.13	2.24	3.03	7.17	−0.79	3.60
Probability	0.99	0.84	0.24	0.79	0.95	1.00	0.31	0.83
Scenario 4: • Penetrance: 80 out of 100 • Effectiveness: 90%								
Utility	5.34	6.65	−0.36	4.75	7.72	5.18	1.16	5.41
Probability	1.00	1.00	0.41	0.89	1.00	0.99	0.76	0.95
Physical disease for which lifestyle changes are possible								
Scenario 5: • Penetrance: 5 out of 100 • Effectiveness: 25%								
Utility	2.03	1.38	−2.19	0.98	2.60	3.76	−1.02	2.20
Probability	0.88	0.80	0.10	0.70	0.93	0.98	0.27	0.80
Scenario 6: • Penetrance: 80 out of 100 • Effectiveness: 25%								
Utility	2.89	4.04	−0.94	2.60	4.85	2.51	0.33	3.06
Probability	0.95	0.98	0.28	0.83	0.99	0.92	0.58	0.88
Scenario 7: • Penetrance: 5 out of 100 • Effectiveness: 75%								
Utility	3.98	2.28	−0.79	2.43	3.60	6.16	0.24	3.73
Probability	0.98	0.91	0.31	0.83	0.97	1.00	0.56	0.89
Scenario 8: • Penetrance: 80 out of 100 • Effectiveness: 75%								
Utility	5.40	6.53	0.58	4.91	7.57	5.16	1.68	5.45
Probability	1.00	1.00	0.64	0.93	1.00	0.99	0.84	0.96

This is of interest because people in the NTTT sample might have undervalued it because they lacked understanding of the attribute or because they did not give enough thought as to how the penetrance probability would influence their decision. Such outcomes were expected based on previous TTT studies[20,22,23] and previously conducted DCE studies in which respondents were not offered any TTT.[42,43] These effects have previously been related to low health literacy or numeracy levels of respondents.[43] Instead of concluding that DCE studies might not be feasible to all subgroups in a population, offering respondents TTT may partly resolve this issue.

In addition, in the NTTT sample, preference reversals were found in one of the classes. In this class, preference was given to receiving information about a disease with low penetrance probability (instead of a high probability) with no preventive opportunities (instead of lifestyle changes, surgery, or medication). Outcomes in this class most likely represent a proportion of respondents who did not understand or did not pay attention to the choice question that was asked (class membership analysis [results not shown] revealed that lower education and lower literate respondents were more likely to belong to this class). They might answer another choice question, namely, “What disease would you prefer to be the outcome of a genetic test?” or “What information would you be able to handle?” This “mistake” was made by a relatively large proportion of the population, given that the probability of belonging to this class in the LCM was 34%. Preference reversals have been previously reported in DCE studies.[17,42,44] Although these divergent outcomes can be explained, they do pose a threat to the validity of the DCE outcomes as they biased the results. Because no reversals were found in the TTT sample, offering respondents TTT might help to increase attention to and understanding of the choice questions.

Applying a TTT approach resulted in a 25% dropout rate; however, response (and dropout) was not selective. Future research using this TTT approach should investigate to what extent dropout rates in the TTT-DCEs are predictable. Based on the outcomes of those studies, recommendations regarding oversampling in a TTT-DCE can be drafted to further increase the validity of this approach. Oversampling seems a highly feasible solution in the current research setting, because it will not greatly affect the research budget, as most surveys use online tools.

Future research should also focus on the length of the TTT period. This period should be balanced between allowing for enough time to properly think about the topic but not too much time that people forget important information or that the study duration is significantly increased.[10] The current study used a 7-day TTT window, whereas previous interview-led TTT-DCE studies showed similar results after a 1-day TTT window.[20,22,23] In addition, research should invest in studying probes that help people to think about or search for additional information, so respondents use the TTT period efficiently. Such research could lead to establishing a framework that describes in which cases a TTT is most helpful and how the TTT period can best be incorporated in a DCE. In addition, further research should explore the actual impact of psychosocial factors (such as attitudes of friends and family members but also personal emotional status) on decision making to better understand to what extent a TTT period captures these elements.

This study was subject to some limitations. First, the study was conducted in a population consisting of healthy research participants aged 50 to 64 years who participated in an extensive health checkup. Study results can therefore not be generalized to younger or older populations. Especially in the older population, one might expect a high benefit of TTT applications, because previous research has shown that older individuals have more difficulty with understanding complicated health information and making ad hoc decisions in DCEs.[45] Additional research should investigate the advantage of this approach in the elderly population. Second, as compared with the general population, people in the current sample were used to making complicated decisions about their health status because they agreed to be part of a research project. Therefore, the effects of using a TTT approach in a DCE study directed at the general population might be more pronounced because they are less familiar with making complex health-related decisions. Third, although the response rate was better than what one can expect for surveys, the response might still be selective. As compared with the average Swedish population, respondents to this study were slightly higher educated. Although higher educated respondents might be more skilled to find and process additional information, the effects of a TTT approach are expected to be even more pronounced among lower educated individuals because the TTT period offers them a larger time window (as compared with standard online surveys) to think about their options and to find additional information or to discuss with peers.. In summary, current findings might underestimate the true value of a TTT approach in DCE studies.

In conclusion, offering respondents TTT in a DCE influences decision making and preferences. Developers of future DCEs regarding complex health-related decisions are advised to consider this TTT approach (ie, mimicking real-life decision making) to enhance the validity of the elicited preferences.

Acknowledgments

The authors would like to thank Semra Ozdemir, PhD, of the Signature Program in Health Services and Systems Research and Lien Centre for Palliative Care at Duke-NUS Medical School in Singapore for her support in the early design of the TTT component of this study. This work was supported by a grant to the project Mind the Risk from the Swedish Foundation for Humanities and Social Sciences (grant PR2013–0123).

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.02.014.

REFERENCES

1. Hensher D, Rose JM, Greene WH. *Applied Choice Analysis*. 2nd ed. Cambridge, UK: Cambridge University Press; 2015.
2. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge University Press; 2000.
3. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care. In: Bateman IJ, ed. *The Economics of Non-market Goods and Resources*. Dordrecht, The Netherlands: Springer; 2008.
4. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 2014;32(9):883–902.
5. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145–172.
6. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving patients' preferences a voice in medical treatment life cycle: the PREFER Public-Private Project. *Patient*. 2017;10(3):263–266.
7. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics*. 2019;37(2):201–226.
8. Veldwijk J. *Discrete Choice Experiments in Public Health*. Utrecht, The Netherlands: Utrecht University; 2015.
9. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis: A Primer*. New York, NY: Cambridge University Press; 2005.
10. Ozdemir S. Improving the validity of stated-preference data in health research: the potential of the time-to-think approach. *Patient*. 2015;8(3):247–255.
11. Lloyd AJ. Threats to the estimation of benefit: are preference elicitation methods accurate? *Health Econ*. 2003;12(5):393–402.
12. Loomis J. What's to know about hypothetical bias in stated preference valuation studies? *J Econ Surv*. 2011;2011(25):2.
13. Jacquemet N, Joule RV, Luchini S, Shogren JF. Do people always pay less than they say? *J Pub Econ Th*. 2011;13(5):857–882.
14. List JA, Gallet CA. What experimental protocol influence disparities between actual and hypothetical stated values? *Environ Res Econ*. 2001;20:241–254.
15. Johnson FR, Hauber BA, Gonzalez JM, Levitan B, Coplan P. *No Time-to-Think About Benefit-Risk Preferences: An Experiment to Test the Validity of Patients' Stated Preferences*. Ithaca, NY: Cornell University; 2010.
16. Determann D, Korfage IJ, Lambooij MS, et al. Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. *PLoS One*. 2014;9(7):e102505.
17. de Bekker-Grob EW, Rose JM, Donkers B, et al. Men's preferences for prostate cancer screening: a discrete choice experiment. *Br J Cancer*. 2013;108(3):533–541.
18. Struik MH, Koster F, Schuit AJ, et al. The preferences of users of electronic medical records in hospitals: quantifying the relative importance of barriers and facilitators of an innovation. *Implement Sci*. 2014;9:69.
19. Whittington D. What have we learned from 20 years of stated preference research is less-developed countries? *Annu Rev Resour Econ*. 2010;2:209–236.
20. Whittington D, Smith K, Okorafor A, et al. Giving respondents time to think in contingent valuation studies: a developing country application. *J Environ Econ Manage*. 1992;22(3):205–225.
21. Marshall DA, Gonzalez JM, MacDonald KV, Johnson FR. Estimating preferences for complex health technologies: lessons learned and implications for personalized medicine. *Value Health*. 2017;20(1):32–39.
22. Cook J, Whittington D, Canh AG, Johnson FR, Nyamete A. Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. *Econ Inquiry*. 2007;45(1):100–114.
23. Cook J, Jeuland M, Maskery B, Whittington D. Giving stated preference respondents "time to think": results from four countries. *Environ Res Econ*. 2011;51:473–496.
24. Viberg Johansson J, Langenskiold S, Segerdahl P, et al. Research participants' preferences for receiving genetic risk information: a discrete choice experiment. *Genet Med*. 2019;21:2381–2389.
25. Marshall DA, Gonzalez JM, Johnson FR, et al. What are people willing to pay for whole-genome sequencing information, and who decides what they receive? *Genet Med*. 2016;18(12):1295–1302.
26. Weymann D, Veenstra DL, Jarvik GP, Regier DA. Patient preferences for massively parallel sequencing genetic testing of colorectal cancer risk: a discrete choice experiment. *Eur J Hum Genet*. 2018;26(9):1257–1265.
27. Bergstrom G, Berglund G, Blomberg A, et al. The Swedish CArdioPulmonary BioImage Study: objectives and design. *J Intern Med*. 2015;278(6):645–659.
28. Meulenkamp TM, Gevers SK, Bovenberg JA, et al. Communication of biobanks' research results: what do (potential) participants want? *Am J Med Genet A*. 2010;152a(10):2482–2492.
29. Jelsig AM, Qvist N, Brusgaard K, Ousager LB. Research participants in NGS studies want to know about incidental findings. *Eur J Hum Genet*. 2015;23(10):1423–1426.

30. Middleton A, Morley KI, Bragin E, et al. Attitudes of nearly 7000 health professionals, genomic researchers and publics toward the return of incidental results from sequencing research. *Eur J Hum Genet*. 2016;24(1):21–29.
31. Allen NL, Karlson EW, Malspeis S, et al. Biobank participants' preferences for disclosure of genetic research results: perspectives from the OurGenes, OurHealth, OurCommunity project. *Mayo Clin Proc*. 2014;89(6):738–746.
32. Bollinger JM, Scott J, Dvoskin R, Kaufman D. Public preferences regarding the return of individual genetic research results: findings from a qualitative focus group study. *Genet Med*. 2012;14(4):451–457.
33. Regier DA, Peacock SJ, Pataky R, et al. Societal preferences for the return of incidental findings from clinical genomic sequencing: a discrete-choice experiment. *CMAJ*. 2015;187(6):E190–E197.
34. Jackson L, Goldsmith L, O'Connor A, Skirton H. Incidental findings in genetic research and clinical diagnostic tests: a systematic review. *Am J Med Genet A*. 2012;158A(12):3159–3167.
35. Murphy J, Scott J, Kaufman D, et al. Public expectations for return of results from large-cohort genetic research. *Am J Bioeth*. 2008;8(11):36–43.
36. Bliemer MCJ, Rose JM, Hensher DA. Efficient stated choice experiments for estimating nested logit models. *Transport Res*. 2009;43:19–35.
37. Wangdahl JM, Martensson LI. The communicative and critical health literacy scale—Swedish version. *Scand J Public Health*. 2014;42:25–31.
38. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a short, 3-item version of the Subjective Numeracy Scale. *Med Decis Making*. 2015;35(8):932–936.
39. Fiebig DG, Keane MP, Louvriere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Market Sci*. 2010;29(3):393–421.
40. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. *Health Econ*. 2005;14(10):1079–1083.
41. Hess S, Rose JM. Can scale and coefficient heterogeneity be separated in random coefficient models? *Transportation*. 2012;39(6):1225–1239.
42. Veldwijk J, Lambooij MS, Kallenberg FG, et al. Preferences for genetic testing for colorectal cancer within a population-based screening program: a discrete choice experiment. *Eur J Hum Genet*. 2016;24(3):361–366.
43. Veldwijk J, van der Heide I, Rademakers J, et al. Preferences for vaccination: does health literacy make a difference? *Med Decis Making*. 2015;35(8):948–958.
44. de Bekker-Grob EW, Bliemer MC, Donkers B, et al. Patients' and urologists' preferences for prostate cancer treatment: a discrete choice experiment. *Br J Cancer*. 2013;109(3):633–640.
45. Veldwijk J, Determann D, Lambooij MS, et al. Exploring how individuals complete the choice tasks in a discrete choice experiment: an interview study. *BMC Med Res Methodol*. 2016;16:45.