ScienceDirect

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

Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models

Esther W. de Bekker-Grob, PhD,[1,2,*] Joffre D. Swait, PhD,[1,2] Habtamu Tilahun Kassahun, PhD,[3] Michiel C.J. Bliemer, PhD,[4] Marcel F. Jonker, PhD,[1,2] Jorien Veldwijk, PhD,[1,2] Karen Cong, MSc,[3] John M. Rose, PhD,[5] Bas Donkers, PhD[1,6]

[1]Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, The Netherlands; [2]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands; [3]Australian Rivers Institute, Griffith University, Brisbane, Queensland, Australia; [4]Business School, University of Sydney, Sydney, New South Wales, Australia; [5]Business School, University of Technology Sydney, Sydney, New South Wales, Australia; [6]Department of Business Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands.

ABSTRACT

Background: Lack of evidence about the external validity of discrete choice experiments (DCEs) is one of the barriers that inhibit greater use of DCEs in healthcare decision making.

Objectives: To determine whether the number of alternatives in a DCE choice task should reflect the actual decision context, and how complex the choice model needs to be to be able to predict real-world healthcare choices.

Methods: Six DCEs were used, which varied in (1) medical condition (involving choices for influenza vaccination or colorectal cancer screening) and (2) the number of alternatives per choice task. For each medical condition, 1200 respondents were randomized to one of the DCE formats. The data were analyzed in a systematic way using random-utility-maximization choice processes.

Results: Irrespective of the number of alternatives per choice task, the choice for influenza vaccination and colorectal cancer screening was correctly predicted by DCE at an aggregate level, if scale and preference heterogeneity were taken into account. At an individual level, 3 alternatives per choice task and the use of a heteroskedastic error component model plus observed preference heterogeneity seemed to be most promising (correctly predicting >93% of choices).

Conclusions: Our study shows that DCEs are able to predict choices—mimicking real-world decisions—if at least scale and preference heterogeneity are taken into account. Patient characteristics (eg, numeracy, decision-making style, and general attitude for and experience with the health intervention) seem to play a crucial role. Further research is needed to determine whether this result remains in other contexts.

Keywords: discrete choice experiment, external validity, healthcare utilization, stated preferences.

VALUE HEALTH. 2019; ■(■):■–■

Introduction

The discrete choice experiment (DCE) technique, originating from mathematical psychology,[1,2] has been introduced in health economics to elicit preferences for health and healthcare.[3] The technique is mainstream in marketing, transport, and environmental economics, where it is used to predict individual and collective choices.[4,5] Its application in healthcare has grown exponentially[6,7] because the method is easy to apply and appears efficient.[8-11] DCEs in health economics are commonly used for valuing health and nonhealth outcomes, investigating trade-offs between health and nonhealth outcomes, and developing priority setting frameworks.[12] Nevertheless, there are also several other areas of applications, for example, DCEs can be used alongside measuring health and nonhealth benefits such as for model parameterization[13] and predicting uptake where there is no information.[14]

Currently, among other barriers, the lack of evidence about the external validity of DCEs inhibits greater use of DCEs in healthcare decision making. External validity of DCE is defined as whether individuals behave in reality as they state in a hypothetical context.[3] Hence the question is "Are healthcare choices

Conflicts of interest: None of the authors have competing interests.

* Address correspondence to: Esther W. de Bekker-Grob, PhD, Erasmus School of Health Policy & Management, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands. Email: debekker-grob@eshpm.eur.nl

https://doi.org/10.1016/j.jval.2019.04.1924

predictable?" To support claims based on DCEs for decision making in clinical or policy contexts, external validity is recognized as an important research question.[6,7,12] Although a DCE may succeed in demonstrating internal validity (eg, how accurately preferences are measured, the extent to which the results are consistent with a priori expectations, and the extent to which the DCE takes account of all things deemed important in the construct's domain[15]),[16-18] this does not guarantee external validity.[16,19,20] Only a handful of studies, mainly outside healthcare, have investigated the external validity of DCEs[21-32]; these studies all focus on final outcomes only.[33] Nevertheless, the investigation of external validity should be much broader.[33] That is, where exactly do discrepancies arise when stated preferences (SPs) derived from DCE do not match revealed preferences (RPs)? Unraveling determinants of external validity allows researchers to improve design, execution, and analysis in DCE studies, providing for more accurate (ie, the degree to which SP values being measured are close to the true RP values) assessments of preferences.

An important starting point in unraveling determinants of external validity of DCEs is to focus on the role of the researcher who must decide on the DCE design and model specification. The aim of this article was to determine whether the number of alternatives in a DCE choice task should reflect the actual decision context, and how complex the choice model specification needs to be to predict the uptake of a healthcare intervention mimicking a real-world decision correctly at an aggregate and individual levels. As we focus on testing designs and models for conducting and analyzing DCE data and their ability to predict the uptake of healthcare choices, we should keep in mind that respondents in SP studies probably have a different amount of information about the healthcare interventions they are evaluating than would likely be the case in RP data. Therefore, in our study, we maintain the choice context constant (ie, both the DCE and "a stated choice task mimicking a real-world healthcare decision" separately asked in the survey are explicitly hypothetical). This means that in our study we did not set out to compare purely SPs to purely RPs, but rather we position our comparison somewhere in between. Thus, we present partly an internal validation and partly an external validation to the extent that hypothetical healthcare choices in the real healthcare context are a stronger test than using a holdout sample from the DCE. Using this approach, we can check whether DCE is able to predict a hypothetical situation representing a real-life choice, which is a minimal requirement for external validity (ie, if DCE fails here, external validity will fail as well).[33]

Methods

Six DCEs were used, which varied in (1) medical condition and (2) the number of alternatives per choice task. Participants were randomized to a study setting. The information we presented about the attributes and their levels was exactly the same as the Dutch national flyer and invitation that participants would receive from their general practitioner or the National Institute for Public Health and the Environment to keep information asymmetry between the hypothetical situation representing the actual decision and the actual decision as small as possible.

Conditions

Two medical conditions were considered: influenza vaccination and colorectal cancer (CRC) screening. These medical fields were chosen because (1) subjects face a real choice because they are not obliged to opt for vaccination or screening, which is vital to test the consistency between stated and actual choices; (2) the number of people facing this decision is large, contributing to the relevance and feasibility of the study; and (3) the decision problems involve different diseases and consequences, which is important for generalizability of the study findings. Approval for the study was obtained from the Medical Ethics Committee, Erasmus MC (MEC-2016-095).

Influenza vaccination

Elderly in the Netherlands, aged 60 years and older, have 2 options regarding influenza vaccination: to opt for vaccination or to opt out. To determine the attributes and levels for the influenza vaccination alternatives, literature,[34-38] interviews with experts in the field of influenza vaccination (n = 4), and 3 focus groups with patients aged 60 years and older from general practices (n = 21) were used (see de Bekker-Grob et al[39] for more details). Five attributes were determined as relevant for subjects to choose to vaccinate or not: vaccination effectiveness (levels: 20%, 40%, 60%, and 80%), risk of severe side effects (levels: 1, 10, and 100 in every million persons), risk of mild side effects (levels: 1, 3, and 5 in every 10 persons), protection duration (levels: 3, 6, and 12 months), and absorption time (levels: 2 and 4 weeks).

CRC screening

In the Netherlands, individuals aged 55 to 75 years have 2 options for CRC screening: to opt into or out of screening. To determine the attributes and levels for the CRC screening alternatives, literature, interviews with experts in the field of CRC screening (n = 3), and 3 focus groups with patients aged 55 to 75 years from general practices (n = 20) were used. Five attributes were determined as relevant for subjects to choose for CRC screening or not: screening effectiveness (levels: 20%, 40%, and 60%), risk of false-negative outcome (levels: 15%, 25%, and 35%), waiting time for fecal occult blood test results (levels: 1, 2, and 3 weeks), waiting time for colonoscopy follow-up test (levels: 2, 4, and 8 weeks), and frequency of the fecal occult blood test (levels: every year, every 2 years, and every 3 years).

DCE Design

Each medical condition had 3 study settings: DCE choice tasks with (1) 2 alternatives (ie, pairs), (2) 3 alternatives (ie, triples), and (3) both (ie, mixed). Of note, the DCE design with 2 alternatives was a better reflection of the actual decision for influenza vaccination and for CRC screening. In each DCE study, subjects were asked to consider all 2 or 3 alternatives in a choice task as realistic alternatives and to choose the alternative that appealed most to them. To maximize the D-efficiency of each DCE design while accommodating substantial respondent heterogeneity, Bayesian heterogeneous DCE designs[40] were used. That is, for all 6 DCE studies (ie, 2 medical conditions times 3 DCE studies [pairs, triples, and mixed]), we generated a heterogeneous DCE design consisting of 10 subdesigns; here, we used the Fortran programming language for computations. Each respondent was assigned 1 randomly selected subdesign containing 16 choice tasks. Together these subdesigns were optimal to estimate a multinomial logit (MNL) model, on the basis of a main-effects utility function with 2-way interactions between the attribute "effectiveness" and the other attributes. The prior preference information (attribute weights) as required for the efficient Bayesian optimization approach was obtained from best guess priors using expert judgment and updated for each of these 6 DCE studies after a pilot run of 100 respondents each.

The choice tasks of the 3 DCE studies of each condition were the following: (1) 16 DCE choice tasks with 2 vaccination/screening alternatives or 1 vaccination/screening alternative and 1 opt-out option; (2) 16 DCE choice tasks with 2 vaccination/

screening alternatives and 1 opt-out option; and (3) 8 DCE choice tasks with 2 vaccination/screening alternatives followed by 8 DCE choice tasks with 2 vaccination/screening alternatives and 1 opt-out option.

Besides these 16 DCE choice tasks per DCE study and questions regarding respondents' characteristics, a choice task was added to the survey mimicking the real-world choice (see Appendix Figures A.1 and A.2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1924 for influenza vaccination and CRC screening, respectively); this additional choice task was placed after the first 8 DCE choice tasks. The choice task representing the real-world choice was kept constant among the DCE studies per condition.

Finally, at the end of the survey, the respondents were asked validated Likert-scale questions related to their decision style,[41] health literacy,[42,43] and numeracy[44,45]; these decision-making skills were of interest on the basis of literature, expert opinions, and focus groups because they all were hypothesized to have an impact on vaccination or screening choices. The questionnaire ended with queries about the complexity and length of the questionnaire.

Sample Size and Recruitment

An online sample of 1419 individuals aged 60 years and older and 1421 individuals aged 55 to 75 years from the Dutch general population, nationally representative in terms of age, sex, education, and geographic region, was recruited for the vaccination and screening condition, respectively, via Survey Sampling International. Calculation of optimal sample sizes for a DCE is complicated because it depends on the true values of the unknown parameters estimated in the choice models.[46] Nevertheless, on the basis of our DCE design and pilot run, and using the sample size calculation of de Bekker-Grob et al,[47] a sample size of 1200 respondents per condition (hence 400 respondents per DCE study) is sufficient to reliably detect preference differences between attribute levels at the 5% significance level.

Choice Modeling

Almost without exception, researchers in the field of health economics have modeled their DCE data within a random utility theory framework.[6,46,48-50] Several random utility specifications have been used to analyze discrete choice data.[6,51,52] Each choice model has its set of features, which should fit best the research question. Because the aim of our study was to determine how complex the choice model (*complexity* is here defined as the capacity to reveal underlying preferences; the more "complex" the model, the more degrees of freedom and the higher the capacity to include certain effects) needs to be to predict a choice mimicking a real-world healthcare decision, we analyzed the DCE data in a systematic way: from a simpler (model A, Table 1) to more and more complex models (models B-D, Table 1). On the basis of common practice in health economics,[6,7] our starting point was the homogeneous preference, homoskedastic MNL (model A, Table 1)[48]:

$$U_{in} = V(X_{in}, \beta) + \varepsilon_{in}. \quad (1)$$

As shown in Equation 1, the latent utility of an alternative i in a choice set C_n (as presented to individual n) is decomposable into 2 additively separable parts: (1) a systematic (explainable) component specified as a function of the attributes of the alternatives $V(X_{in}, \beta)$ and (2) a random (unexplainable) component ε_{in} representing stochastic variation in choices. The MNL model has 3 key properties: (1) error terms are assumed independent and identically extreme value type I distributed across observations (IID); (2) independence of irrelevant alternatives, resulting from (1); and (3) no unattributable preference heterogeneity. Such assumptions may be restrictive in describing human behavior, perhaps compromising the external validity of DCE results. Therefore, we started by first relaxing the IID assumption (heteroskedastic multinomial logit [HMNL]; model B, Table 1) followed by relaxing preference homogeneity with known (up to 21 observed) subject characteristic sources (HMNL+; model C, Table 1) and by taking randomness of the alternative specific constant (ASC) parameter capturing systematic preferences toward opting out into account (HMNL++; model D, Table 1). That is, in model D we aimed to explain heterogeneity by including subject characteristics *and* using choice responses to obtain individual conditional parameters (see later) to explain heterogeneity regarding the ASC (ie, α_n). In this way, we are able to distinguish explained heterogeneity from unexplained heterogeneity. Hence, the following choice processes with scale and/or preference heterogeneity and/or random opt-out utility were used (see Equations 2-5):

Table 1. Systematic choice modeling approach.

Characteristics	Model			
	A	B	C	D
1. Random-utility-maximization decision rule	x	x	x	x
2. Scale heterogeneity		x	x	x
3. Systematic preference heterogeneity			x	x
4. Random opt-out utility				x

$$U(\text{opt}-\text{out})_{in} = \alpha_n + \varepsilon_{in}$$

$$U(\text{opt}-\text{in})_{in} = \beta X_i + \delta Z_{1n} X_i + \varepsilon_{in} \quad (2)$$

where

$$\alpha_n = \overline{\alpha} + \theta Z_{1n} + \eta_n, \quad (3)$$

$$\eta_n = N\left(0, (\sigma_\alpha)^2\right),$$

$$\varepsilon_{in} = \text{HEV}(\mu_n), \quad (4)$$

$$\mu_n = \exp(\gamma Z_{2n}). \quad (5)$$

The quantity α_n is the ASC for opting out for vaccination/screening compared with opting in, Z_1 and Z_2 are 2 sets of subject characteristics, η_n is a normally distributed random component in α_n, and HEV is a heteroskedastic extreme value distribution with variance parameter μ_n.[53,54] To use terminology common to discrete choice models, our full specification is a heteroskedastic error component model.[55,56] For each DCE data set, we used this 4-step approach (models A-D) to determine the optimal utility function using the Pythonbiogeme software (Michiel Bierlaire, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland)[57,58] and taking the best model fits into account on the basis of the Bayesian information criterion.

Is Patient Choice Predictable?

For each DCE study, we first determined for the fixed choice task mimicking the real-world choice which proportion of the sample opted for vaccination or screening (ie, the observed uptake). Second, using the parameter values estimated from the DCE choice tasks, we determined for each DCE study the predicted vaccination/screening uptake on the basis of the 4 best-fitting choice models (ie, models A-D, Table 1). Finally, we determined to what extent the predicted uptake was in agreement with the observed uptake at an aggregate and individual levels using probability rules, mean values, and 95% confidence intervals, whereas at an individual level for model D conditional parameter estimates were taken into account as well. Regarding the latter, using the coefficients of model D, the Pythonbiogeme software, the conditional parameter approach of Train[59] and Revelt and Train,[60] and Excel (Hess[61]), we were able to determine the ASCs per individual (ASC_*i*; hence, α_n). These ASC_*i*'s were added to the data set as a variable, because if individuals have systematically different preferences, which are unrelated to observed characteristics, ignoring to address it can bias the estimates of the preference weights. The utility function of model D, incorporating the ASC_*i*'s, was used to determine the utility weight for each individual for influenza vaccination in the choice task mimicking the real-world decision. Then, like models A to C, we first simulated the probabilities for each individual that he or she would opt for influenza vaccination. Second, we summarized the probabilities for influenza vaccination for respondents who opted for vaccination in the choice task mimicking the real-world decision. Third, we summarized the probabilities for no vaccination for respondents who opted out of influenza vaccination in the choice task mimicking the real-world decision. Finally, these probability scores were summarized, divided by the sample size, and multiplied by 100 to obtain the percentage of correctly predicted choices at an individual level.[62] The same procedure was followed for the CRC screening condition.

Results

Respondents

In total, 1429 and 1421 subjects were recruited to participate in the DCE studies concerning influenza vaccination and CRC screening, respectively. This resulted in 1261 (89%) and 1219 (86%) completes. In both conditions, there were no substantial differences between the 3 DCE designs regarding dropout rates or perceived burden (see Table 2).

Respondents of the vaccination condition had a mean age of 66 ± 5 years, about 56% were men, and one-third had a lower educational level (see Appendix Table A.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1924). Approximately 75% of the respondents reported that they were in good health, 27% had experienced influenza (symptoms) last year, and 30% of the respondents mentioned that they had never been vaccinated against influenza. There were no significant differences between the 3 samples in respondents' characteristics, except for the variable sex.

Table 2. Number of completes, dropout rates, and perceived burden for the influenza vaccination and CRC screening DCE surveys.

Issues	Pairs		Triples		Mixed	
	n	%	n	%	n	%
Influenza vaccination						
Responsiveness						
Completes	423		418		420	
Dropouts	52		57		59	
Dropout rate		10.9		12.0		12.3
Difficulty filling the questionnaire (yes)		2.6		2.9		2.9
Very easy	104		77		89	
Easy	196		217		225	
Neutral	112		112		94	
Difficult	10		11		11	
Very difficult	1		1		1	
Length of the questionnaire (good)		83.0		79.7		85.5
Too long	72		85		69	
Not too long, not too short	351		333		359	
Too short	0		0		0	
CRC screening						
Responsiveness						
Completes	406		406		407	
Dropouts	57		79		66	
Dropout rate		12.3		16.3		14.0
Difficulty filling the questionnaire (yes)		8.9		9.4		7.9
Very easy	40		27		39	
Easy	140		166		163	
Neutral	190		175		173	
Difficult	31		35		28	
Very difficult	5		3		4	
Length of the questionnaire (good)		82.0		81.8		81.6
Too long	72		74		75	
Not too long, not too short	333		332		332	
Too short	1		0		0	

CRC indicates colorectal cancer; DCE, discrete choice experiment.

Table 3. DCE (triples) results: influenza vaccination survey.

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value
ASC									
	No vaccination	0.974	<.010*	0.522	<.010*	2.960	<.010*	5.630	<.010*
Attributes (main effects)									
Scaled (/10)	Effectiveness	0.146	<.010*	0.059	<.010*	0.089	.010*	0.148	<.010*
	Serious side effects								
	1/1.000.000	0.283		0.120		0.118		0.391	
	10/1.000.000	0.227	<.010*	0.107	.020†	0.247	.040†	0.193	.110
	100/1.000.000	−0.510	<.010*	−0.227	.010*	−0.365	.010*	−0.584	<.010*
	Mild side effects								
	1/10	0.078		0.010		0.283		0.247	
	3/10	−0.011		0.011	.750	−0.342	<.010*	−0.262	.030†
	5/10	−0.067	.330	−0.020	.560	0.059	.600	0.015	.900
	Protection duration								
	3 mo	−0.271		−0.120		−0.029		−0.170	
	6 mo	0.165	.020†	0.063	.110	0.249	.010*	0.347	<.010*
	12 mo	0.106	.120	0.057	.130	−0.220	.040†	−0.177	.120
	Waiting time								
	2 wk	0.013		−0.019		−0.037		−0.036	
	4 wk	−0.013	.780	0.019	.430	0.037	<.010*	0.036	<.010*
Two-way interactions									
Scaled (/10)	eff × serious10	−0.315	.010*	−0.143	.050†	−0.261	.070‡	−0.155	.280
Scaled (/10)	eff × serious100	0.192	.150	0.144	.080‡	0.046	.780	−0.127	.440
Scaled (/100)	eff × mild3	0.530	.660	0.018	.970	1.630	.280	1.360	.370
Scaled (/100)	eff × mild5	−0.242	.840	−0.110	.850	−0.619	.680	−1.370	.370
Scaled (/10)	eff × dur6	−0.246	.040†	−0.080	.200	−0.172	.250	−0.209	.170
Scaled (/10)	eff × dur12	0.237	.050†	0.042	.500	0.407	.010*	0.506	<.010*
Scaled (/100)	eff × wait4	0.059	.470	−0.021	.610	0.022	.710	−0.038	.550
Scale heterogeneity									
	Good health literacy	–		0.362	<.010*	0.030	.510	−0.015	.860
	Good numeracy	–		0.264	.020†	0.032	.490	0.138	.110
	Male	–		−0.488	<.010*	0.237	<.010*	0.032	.750
	Age > 65 y	–		−0.344	<.010*	−0.006	.910	−0.211	.010*
	Family impact on decision	–		−1.790	.070‡	−0.118	.230	−0.037	.790
	GP visit last month	–		−1.500	<.010*	−0.184	<.010*	−0.006	.940
	Good health	–		1.110	<.010*	−0.287	<.010*	0.037	.740
Systematic preference heterogeneity									
	Age > 65 y × constant "no vacc"	–		–		0.545	<.010*	2.620	.010*
	Age > 65 y × eff	–		–		−0.042	.080‡	−0.023	.410
	Attitude for × constant "no vacc"	–		–		−1.810	<.010*	−5.74	<.010*
	Attitude for × dur6	–		–		−0.084	.290	−0.125	.120
	Attitude for × dur12	–		–		0.426	<.010*	0.394	<.010*
	Attitude for × eff	–		–		0.235	<.010*	0.164	<.010*

continued on next page

Table 3. Continued

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value
	Attitude for × serious10	–		–		0.012	.870	0.016	.820
	Attitude for × serious100	–		–		−0.189	.050†	−0.041	.680
	Attitude for × wait4	–		–		−0.149	<.010*	−0.134	.010*
	No disease × constant "no vacc"	–		–		−1.020	<.010*	−1.580	.130
	No disease × mild3	–		–		0.217	<.010*	0.170	.020†
	No disease × mild5	–		–		−0.028	.700	−0.008	.920
	Deliberative DM style × eff	–		–		0.043	.020†	0.063	.040†
	Deliberative DM style × mild3	–		–		0.159	.060‡	0.157	.060†
	Deliberative DM style × mild5	–		–		−0.122	.160	−0.130	.130
	High education × eff	–		–		0.078	<.010*	0.097	<.010*
	High education × dur6	–		–		−0.160	.030†	−0.173	.020†
	High education × dur12	–		–		−0.001	.990	0.005	.950
	Impact family × constant "no vacc"	–		–		−0.423	.040†	−2.660	.140
	Flu symptoms last year × "no vacc"	–		–		0.289	.020†	0.416	.710
	Last month GP visit × mild3	–		–		0.183	.020†	0.128	.090‡
	Last month GP visit × mild5	–		–		−0.125	.110	−0.112	.150
	Last month GP visit × serious10	–		–		0.016	.830	−0.001	.990
	Last month GP visit × serious100	–		–		−0.217	.020†	−0.134	.170
	Good health × constant "no vacc"	–		–		1.330	<.010*	2.450	.040†
	Good health × eff	–		–		0.103	<.010*	0.069	.060‡
	Good health × serious10	–		–		0.085	.230	0.073	.330
	Good health × serious100	–		–		−0.212	.020†	−0.072	.550
	Good health literacy × constant "no vacc"	–		–		−0.384	.020†	−0.173	.840
	Good health literacy × eff	–		–		−0.089	<.010*	−0.087	<.010*
	Good numeracy × constant "no vacc"	–		–		1.110	<.010*	1.980	.030†
	Good numeracy × eff	–		–		0.127	<.010*	0.134	<.010*
	Male × constant "no vacc"	–		–		−1.090	<.010*	−1.670	.070†
	Male × eff	–		–		−1.010	<.010*	−0.098	<.010*
	Male × serious10	–		–		−0.171	.010*	−0.175	.010*
	Male × serious100	–		–		0.260	<.010*	0.165	.100‡
	Vacc last year × constant "no vacc"	–		–		−3.060	<.010*	−8.980	<.010*
	Vacc last year × serious10	–		–		0.011	.890	0.016	.840
	Vacc last year × serious100	–		–		−0.381	<.010*	−0.261	.010*
	Vacc last year × dur6	–		–		−0.037	.660	−0.086	.310

continued on next page

Table 3. Continued

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value
	Vacc last year × dur12	–		–		0.350	<.010*	0.301	<.010*
	No side effects × constant "no vacc"	–		–		−1.150	<.010*	−3.700	<.010*
	No side effects × mild3	–		–		0.054	.450	0.031	.680
	No side effects × mild5	–		–		−0.178	.020†	−0.113	.140
	Flu although vacc × constant "no vacc"	–		–		0.416	<.010*	1.740	.180
Random opt-out utility (SD of ASC)									
		–		–		–		6.400	<0.01*
Goodness of fit									
	LL	−7106		−7060		−4713		−3459	
	Number of free parameters	16		23		68		69	
	AIC	2.130		2.118		1.430		1.055	
	BIC	2.139		2.132		1.471		1.097	
	Respondents	418		418		418		418	

AIC indicates Akaike information criterion; ASC, alternative specific constant; BIC, Bayesian information criterion; DCE, discrete choice experiment; DM, decision-making; GP, general practitioner; HMNL, heteroskedastic model; LL, log likelihood; MNL, multinomial model; SD, standard deviation.
*Significant at the 1% level.
†Significant at the 5% level.
‡Significant at the 10% level.

Respondents of the screening condition had a mean age of 63 ± 5 years, about 50% were men, and one-third had a lower educational level (see Appendix Table A.2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1924). Approximately 73% of the respondents reported that they were in good health, 13% suffered from cancer, and 73% mentioned that they have experience with screening. There were no significant differences in respondents' characteristics across the 3 samples.

Model Fit and Parameters

In the context of influenza vaccination data and focusing on the sample of respondents who received 3 alternatives per choice task (ie, triples), for all 4 models, the directions of the coefficients of the vaccination attributes were consistent with our a priori hypotheses, which implies theoretical validity of the different models (Table 3). The more complex/sophisticated the model, the better the model fit. There is a very large difference in model fit between the models that took preference heterogeneity into account versus the simpler models. Looking at the model with the best model fit (model D), all vaccination attributes influenced patients' decision behavior for influenza vaccination. There is evidence of substantial scale and preference heterogeneity. Regarding scale heterogeneity, older respondents were significantly less consistent (P=.010) in their choices (ie, more affected by random variation in utility[29]). Regarding preference heterogeneity, 11 respondent characteristics were observed that explained the preference heterogeneity found (P<.050): age, sex, education, health literacy, numeracy, decision-making style, health state, having a chronic disease, general attitude to vaccination, having been vaccinated last year, and experience with side effects of vaccination. Model D was also the best-fit model for the vaccination sample of respondents who filled in the pairs or the mixed choice tasks (see Appendix Tables A.3 and A.4, respectively, in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1924).

Respondents' characteristics that explained scale and preference heterogeneity differed between the 3 DCE vaccination samples. Nevertheless, 7 characteristics explained part of the observed scale and/or preference heterogeneity in all 3 DCE vaccination samples: sex, health literacy, numeracy, decision-making style, general attitude to vaccination, having a disease, and having been vaccinated last year.

In the context of CRC screening data, again focusing on the sample of respondents who received the DCE with triples, for all 4 models the directions of the coefficients of the screening attributes showed the expected signs and most of them were statistically significant (Table 4). Also in this sample, model D had the best model fit. Again substantial scale and preference heterogeneity were found.

Regarding scale heterogeneity, respondents who had a more deliberative decision style, lived alone, or had a higher level of numeracy were more consistent in their choices (P<.010). Regarding preference heterogeneity, 12 respondent characteristics were observed that explained the preference heterogeneity found (P<.050): age, sex, education, health literacy, numeracy, decision-making style, health state, general attitude toward screening, living alone, CRC history in family, screening experience, and having consulted a general practitioner last month. Model D was also the model with the best model fit for the screening sample of

Table 4. DCE (triples) results: CRC screening survey.

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value	Coefficient value	*P* value
ASC									
	No CRC screening	−2.710	<.010*	−2.620	<.010*	0.517	.230	−0.407	.910
Attributes (main effects)									
Scaled (/10)	Effectiveness	0.020	<.010*	0.018	<.010*	0.008	.200	0.016	.040
	False negative	−0.048	<.010*	−0.052	<.010*	0.001	.940	−0.022	.150
	Frequency								
	Every year	0.133		0.127		−0.132		−0.176	
	Every 2 y	0.144	.020†	0.214	<.010*	0.335	<.010*	0.342	<.010*
	Every 3 y	−0.277	<.010*	−0.341	<.010*	−0.203	.090‡	−0.166	.180
	Waiting time diagn test								
	1 wk	−0.033		−0.180		−0.116		−0.096	
	2 wk	0.034	.580	0.241	<.010*	0.174	.010*	0.194	<.010*
	3 wk	−0.001	.990	−0.061	<.010*	−0.058	.400	−0.099	.150
	Waiting time f-up test								
	2 wk	−0.091		0.131		0.239		0.218	
	4 wk	0.119	.060‡	0.036	.400	0.019	.740	0.048	.360
	8 wk	−0.028	.660	−0.167	<.010*	−0.258	<.010*	−0.265	<.010*
Two-way interactions									
Scaled (/10)	eff × fneg	0.001	.640	0.001	.550	0.001	.510	0.002	.180
Scaled (/10)	eff × freq2	0.008	.590	−0.021	.030†	−0.032	.010*	−0.032	.010*
Scaled (/10)	eff × freq3	−0.019	.230	0.026	.010*	0.047	<.010*	0.049	<.010*
Scaled (/100)	eff × waitdiag2	0.003	.850	−0.009	.350	−0.009	.430	−0.012	.260
Scaled (/100)	eff × waitdiag3	0.004	.800	0.018	.080‡	0.015	.220	0.020	.080‡
Scaled (/10)	eff × wait f-up4	−0.185	.220	−0.017	.860	0.062	.600	0.005	.970
Scaled (/10)	eff × wait f-up8	−0.281	.050†	0.085	.370	0.054	.640	0.083	.440
Scale heterogeneity									
	Age > 65 y	–		−0.356	<.010*	0.633	<.010*	0.743	.110
	Did not have (had) cancer	–		0.285	<.010*	−0.090	<.010*	−0.122	.310
	Rather deliberative decision making	–		0.403	<.010*	0.493	<.010*	0.352	<.010*
	Bad experience	–		−2.410	<.010*	−0.333	.090‡	−0.541	.170
	Health literacy	–		−0.246	<.010*	0.104	.090‡	0.059	.510
	Living alone	–		0.337	<.010*	0.562	<.010*	0.540	<.010*
	Male	–		0.253	<.010*	−0.456	<.010*	−0.213	.040
	Good numeracy	–		0.279	<.010*	−0.575	<.010*	−0.350	<.010*
	Did not have screening experience	–		−1.490	<.010*	−0.128	.070‡	−0.132	.220
Systematic preference heterogeneity									
	Age > 65 y × constant "no screening"	–		–		1.730	<.010*	1.690	.400
	Age > 65 y × eff	–		–		−0.021	<.010*	−0.026	<.010*
	Age > 65 y × fneg	–		–		0.033	<.010*	0.040	<.010*
	Age > 65 y × freq2	–		–		−0.111	.020†	−0.132	.010*
	Age > 65 y × freq3	–		–		0.149	.010*	0.165	.010*

continued on next page

Table 4. Continued

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
	Age > 65 y × waitdiagn2	–		–		−0.033	.440	−0.051	.220
	Age >65 y × waitdiagn3	–		–		−0.130	.010*	−0.100	.040†
	Attitude for × constant "no screening"	–		–		−3.760	<.010*	−15.500	<.010*
	Attitude for × eff	–		–		0.020	<.010*	0.014	.020†
	Attitude for × fneg	–		–		−0.047	<.010*	−0.035	<.010*
	Attitude for × freq2	–		–		0.067	.430	0.065	.440
	Attitude for × freq3	–		–		−0.449	<.010*	−0.503	<.010*
	No cancer × constant "no screening"	–		–		−1.060	<.010*	−4.740	.070‡
	No cancer × fneg	–		–		−0.011	.100‡	−0.009	.160
	CRC in family × eff	–		–		−0.005	.090‡	−0.007	.060‡
	CRC in family × fneg	–		–		0.021	<.010*	0.020	<.010*
	CRC in family × waitdiagn2	–		–		−0.081	.060‡	−0.074	.080‡
	CRC in family × waitdiagn3	–		–		0.092	.050†	0.078	.100‡
	Deliberative DM style × constant "no screening"	–		–		−0.408	.070‡	−4.910	.050†
	Deliberative DM style × fneg	–		–		−0.013	.020†	−0.015	.020†
	Deliberative DM style × freq2	–		–		−0.076	.080‡	−0.075	.070‡
	Deliberative DM style × freq3	–		–		0.080	.090‡	0.085	.070‡
	High education × constant "no screening"	–		–		−0.859	<.010*	−2.990	.190
	High education × fneg	–		–		−0.011	.040†	−0.012	.020†
	Bad experience × fneg	–		–		−0.032	<.010*	−0.036	.300
	Bad experience × freq2	–		–		0.366	.040†	0.539	.090‡
	Bad experience × freq3	–		–		−0.255	.170	−0.179	.480
	Last month GP visit × eff	–		–		−0.008	.010*	−0.009	.010*
	Last month GP visit × fneg	–		–		0.015	<.010*	0.014	.020†
	Last month GP visit × freq2	–		–		0.007	.870	0.015	.710
	Last month GP visit × freq3	–		–		−0.136	.010*	−0.123	.010*
	Last month GP visit × waitdiagn2	–		–		−0.100	.010*	−0.086	.030†
	Last month GP visit × waitdiagn3	–		–		0.064	.140	0.056	.180
	Good health × constant "no screening"	–		–		−0.636	<.010*	0.144	.950
	Good health × fneg	–		–		−0.011	.060‡	−0.009	.110
	Good health × freq2	–		–		0.086	.070‡	0.090	.050†
	Good health × freq3	–		–		−0.061	.240	−0.060	.220
	Good health literacy × constant "no screening"	–		–		0.628	<.010*	0.677	.730
	Good health literacy × fneg	–		–		0.012	.020†	0.011	.050†
	Good health literacy × freq2	–		–		−0.103	.020†	−0.091	.020†

continued on next page

Table 4. Continued

Triples	Utility function	Model A (MNL model)		Model B (HMNL model)		Model C (HMNL model + systematic preference heterogeneity)		Model D (HMNL model + systematic preference heterogeneity + random opt-out utility)	
		Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
	Good health literacy × freq3	–		–		0.057	.220	0.066	.130
	Last month hospital visit × eff	–		–		0.005	.080†	0.002	.440
	Living alone × constant "no screening"	–		–		0.924	<.010*	1.110	.580
	Living alone × freq2	–		–		−0.064	.140	−0.073	.070‡
	Living alone × freq3	–		–		0.230	<.010*	0.206	<.010*
	Living alone × wait f-up4	–		–		−0.037	.340	−0.047	.190
	Living alone × wait f-up8	–		–		0.076	.050†	0.080	.030†
	Male × constant "no screening"	–		–		−1.300	<.010*	−1.360	.530
	Male × fneg	–		–		−0.013	.060‡	−0.006	.330
	Male × freq2	–		–		0.100	.080‡	0.073	.130
	Male × freq3	–		–		−0.224	<.010*	−0.151	.020†
	Male × waitdiagn2	–		–		0.014	.780	−0.010	.790
	Male × waitdiagn3	–		–		0.107	.040†	0.057	.170
	Good numeracy × constant "no screening"	–		–		−1.720	<.010*	−0.030	.990
	Good numeracy × eff	–		–		0.028	<.010*	0.024	<.010*
	Good numeracy × fneg	–		–		−0.064	<.010*	−0.052	<.010*
	No screening experience × constant "no screening"	–		–		1.990	<.010*	7.180	.010*
	No screening experience × freq2	–		–		−0.119	.080‡	−0.123	.060‡
	No screening experience × freq3	–		–		−0.031	.660	−0.056	.430
	No screening experience × wait f-up4	–		–		0.017	.760	-0.001	.990
	No screening experience × wait f-up8	–		–		0.236	<.010*	0.184	<.010*
Random opt-out utility (SD of ASC)									
		–		–		–		10.200	<.010*
Goodness of fit									
	LL	−5614		−5265		−4778		−4084	
	Number of free parameters	16		25		86		87	
	AIC	1.734		1.629		1.497		1.284	
	BIC	1.743		1.644		1.551		1.338	
	Respondents	406		406		406		406	

AIC indicates Akaike information criterion; ASC, alternative specific constant; BIC, Bayesian information criterion; CRC, colorectal cancer; DCE, discrete choice experiment; DM, decision-making; GP, general practitioner; HMNL, heteroskedastic model; LL, log likelihood; MNL, multinomial model; SD, standard deviation.
*Significant at the 1% level.
†Significant at the 5% level.
‡Significant at the 10% level.

respondents who filled in the pairs and the triples (see Appendix Tables A.5 and A.6, respectively, in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1924). Respondent characteristics that explained scale and preference heterogeneity also differed between the 3 DCE screening samples. Nevertheless, 8 respondent characteristics explained part of the observed scale and/or preference heterogeneity in all 3 DCE screening samples: age, sex, numeracy, decision-making style, general attitude toward

Table 5. DCE model fit and prediction results based on influenza vaccination.

Model fit and prediction results	Pairs				Triples				Mixed			
	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)
Goodness of fit												
LL	−4357	−4329	−4087	−4086	−7106	−7060	−5149	−3614	−5699	−5682	−4793	−3949
No. of free parameters	14	21	51	52	16	23	59	60	14	19	54	55
AIC	1.292	1.286	1.223	1.223	2.130	2.118	1.558	1.099	1.700	1.697	1.443	1.192
BIC	1.300	1.298	1.253	1.254	2.139	2.132	1.593	1.135	1.709	1.708	1.475	1.225
Respondents	423	423	423	423	418	418	418	418	420	420	420	420
Vaccination uptake												
Observed*	64.5%	64.5%	64.5%	64.5%	61.2%	61.2%	61.2%	61.2%	66.2%	66.2%	66.2%	66.2%
Predicted mean	73.3%	76.6%	66.2%	65.9%	56.8%	48.8%	58.4%	61.7%	58.9%	55.7%	60.9%	65.0%
Delta	8.8%	12.1%	1.7%	1.4%	−4.4%	−12.4%	−2.8%	0.5%	−7.3%	−10.5%	−5.3%	−1.2%
Lower bound CI	68.8%	72.2%	61.4%	61.2%	51.8%	43.9%	53.5%	56.9%	53.9%	50.8%	56.1%	60.2%
Upper bound CI	77.4%	80.6%	70.7%	70.4%	61.5%	53.7%	63.1%	66.4%	63.6%	60.5%	65.6%	69.6%
Proportion of choices that were predicted correctly at an individual level	56.8%	59.6%	74.6%	81.6%	51.5%	49.9%	81.7%	93.6%	52.9%	51.7%	78.6%	88.7%

AIC indicates Akaike information criterion; BIC, Bayesian information criterion; CI, confidence interval; DCE, discrete choice experiment; HMNL, heteroskedastic model; HMNL+, heteroskedastic model with preference heterogeneity; HMNL++, heteroskedastic model with preference heterogeneity and random intercept; LL, log likelihood; MNL, multinomial model.
*Vaccination uptake observed is the proportion of respondents who opted for influenza vaccination in the fixed-choice task mimicking real-life choice task.

CRC screening, screening experience, living alone, and health state.

Are Healthcare Choices Predictable?

Irrespective of the number of alternatives per choice task, the choice to opt for influenza vaccination and CRC screening was correctly predicted by the DCE at an aggregate level if scale, preference heterogeneity, and a random opt-out utility were taken into account (Table 5 and Table 6 respectively; model D). A similar phenomenon was seen for 4 of 6 DCE samples, if only scale and observed preference heterogeneity were taken into account (model C). For both medical conditions we found that a better model fit did not automatically mean better prediction (eg, HMNL vs MNL).

At an individual level, the choice for vaccination was predicted best using a heteroskedastic error component model that took into account observed preference heterogeneity through subject

Table 6. DCE model fit and prediction results based on colorectal survey.

Model fit and prediction results	Pairs				Triples				Mixed			
	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)	Model A (MNL)	Model B (HMNL)	Model C (HMNL+)	Model D (HMNL++)
Goodness of fit												
LL	−4357	−4329	−4087	−4086	−7106	−7060	−5149	−3614	−5699	−5682	−4793	−3949
No. of free parameters	14	21	51	52	16	23	59	60	14	19	54	55
AIC	1.292	1.286	1.223	1.223	2.130	2.118	1.558	1.099	1.700	1.697	1.443	1.192
BIC	1.300	1.298	1.253	1.254	2.139	2.132	1.593	1.135	1.709	1.708	1.475	1.225
Respondents	423	423	423	423	418	418	418	418	420	420	420	420
CRC screening uptake												
Observed*	92.9%	92.9%	92.9%	92.9%	92.9%	92.9%	92.9%	92.9%	91.9%	91.9%	91.9%	91.9%
Predicted mean	97.4%	96.6%	95.2%	95.0%	84.9%	86.5%	88.0%	90.6%	90.1%	90.4%	91.6%	93.0%
Delta	4.5%	3.7%	2.3%	1.4%	−8.0%	−6.4%	−4.9%	−2.3%	−1.8%	−1.5%	−0.3%	−1.2%
Lower bound CI	95.2%	94.3%	92.8%	92.5%	81.1%	82.7%	84.4%	87.4%	86.9%	87.1%	88.5%	90.2%
Upper bound CI	98.6%	98.1%	97.2%	97.0%	88.3%	89.6%	90.9%	93.3%	92.9%	93.1%	94.1%	95.4%
Proportion of choices that were predicted correctly at an individual level	80.6%	89.9%	90.5%	91.2%	79.9%	83.1%	87.9%	97.1%	83.6%	84.8%	89.0%	93.9%

AIC indicates Akaike information criterion; BIC, Bayesian information criterion; CI, confidence interval; CRC, colorectal cancer; DCE, discrete choice experiment; HMNL, heteroskedastic model; HMNL+, heteroskedastic model with preference heterogeneity; HMNL++, heteroskedastic model with preference heterogeneity and random intercept; LL, log likelihood; MNL, multinomial model.
*CRC screening uptake observed is the proportion of respondents who opted for CRC screening in the fixed-choice task mimicking real-life choice task Appendix.

characteristics and unknown subject characteristic sources that systematically affect the preference for opting out (model D): 81.6%, 93.6%, and 88.7% of the individuals' choices were correctly predicted, using a DCE with 2 alternatives or 3 alternatives per choice task, or both, respectively.

At an individual level, the choice for screening was also predicted best using model D: 91.2%, 97.1%, and 93.9% of the individuals' choices were correctly predicted, using pairs, triples, and mixed choice tasks, respectively. Overall, looking at the results of both surveys, using the 3 alternatives per choice task, a heteroskedastic error component model with observed preference heterogeneity, and conditional parameter estimates for the unobserved preference heterogeneity for opting out seemed to correctly predict choices mimicking real-world decisions in more than 93% of the respondents at an individual level. As we are getting high prediction success statistics with this rather noncomplex model, our objective is attained; there is no clear rationale for using more complex (ie, advanced) models here.

Discussion

This study showed that irrespective of the number of alternatives per choice task over the range tested, the choice mimicking a real-world decision to opt for influenza vaccination and CRC screening was correctly predicted by a DCE-based model at an aggregate level, if scale and preference heterogeneity were taken into account. At an individual level, using the 3 alternatives per choice task and a heteroskedastic error component model seemed to be most promising, correctly predicting in 93.6% and 97.1% of the cases for vaccination and screening, respectively. Five respondent characteristics consistently explained a part of the observed scale and/or preference heterogeneity: sex, numeracy skill, decision-making style, general attitude toward the health intervention of interest, and experience with the health intervention of interest. For the models and designs we used, our study showed that more than 93% of choices were correctly predicted at an individual level, giving a degree of confidence in the results and external validity of SP studies.

A healthcare DCE study of Mohammadi et al[63] found that their best model correctly predicted at an individual level in 83% of the participants, which is somewhat lower than our findings. The same study however showed that individual-specific coefficients reflected respondents' actual choices more closely compared with aggregate-level estimates, which agrees with our outcomes. Wright et al[64] mentioned that it is likely that if scale heterogeneity is unaddressed in a healthcare DCE, the results of such a study might be misleading. Our study confirmed that this is indeed the case, but additionally our results showed that if systematic preference heterogeneity is ignored, the results of healthcare DCE studies might be misleading as well.

Although 2 or 3 alternatives per DCE choice task, or both, all predicted well at an aggregate level, we recommend the use of the 3 alternatives per DCE choice task design (at least if an opt-out alternative is included). Literature outside healthcare already shows that more robust choice models can be constructed from the 3 alternatives per choice task design than from the 2 alternatives per choice task design.[65] In addition, our results showed that the 3 alternatives per choice task DCEs did not differ in dropout rates or perceived burden compared with our other DCE samples. Nevertheless, we found a slightly better prediction at an individual level in the case of the 3 alternatives per choice task design, probably because of a richer data set. Note that our results are conditional on the number of attributes used in our study; decisions involving more attributes should view our recommendations as indicative.

Our study also has several limitations. First, there is the issue of generalizability of our results. Although the results were based on 2 unlabeled DCE designs in different disease areas, further research is needed to determine whether our results hold in labeled DCE designs and other disease contexts. Second, because of the systematic approach we used starting from common practice in health economics, other choice models (eg, latent class model and mixed logit model) and ensemble analyses were not investigated. Further research is therefore recommended, although researchers should be cautious about overfitting their data. In addition, further research is recommended to see whether DCE holds its potential when predicting real-world decisions, because issues such as hypothetical bias and difference in amount of information may play an important role.

Conclusions

Our study shows that DCE models that take into account both scale and preference heterogeneity were better able to predict choices mimicking real-world decisions and performed best when 3 alternatives were presented to respondents. Patient characteristics such as sex, numeracy, decision-making style, general attitude toward the health intervention of interest, and experience with the health intervention of interest seem to play a crucial role in predicting healthcare choices. Further research is needed to determine whether this result remains in other contexts.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. We also want to thank the general practitioners Patrick Bindels, Sylvia Buis and Jan Huisman for their great help with organizing the focus groups.

Source of Financial Support

This study was supported by a grant from The Netherlands Organisation for Scientific Research (NWO-Talent-Scheme-Veni-Grant No. 451-15-039). The funder had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the manuscript for publication.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.04.1924.

REFERENCES

1. Thurstone L. A law of comparative judgment. *Psychol Rev*. 1927;34:273–286.
2. Luce RD, Tukey JW. Simultaneous conjoint-measurement—a new type of fundamental measurement. *J Math Psychol*. 1964;1(1):1–27.
3. Ryan M. Discrete choice experiments in health care. *BMJ*. 2004;328(7436):360–361.
4. Mahieu PA, Andersson H, Beaumais O, Crastes Sourd R, Hess S, Wolff FC. Stated preferences: a unique database composed of 1,657 recent published articles in journals related to agriculture, environment or health. *Rev Agric Food Environ Stud*. 2017;98(3):201–220.
5. Bliemer MCJ, Rose JM. Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transp Res Part A Policy Pract*. 2011;45(1):63–79.
6. 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.
7. Clark M, 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.
8. Rowen D, Mulhern B, Stevens K, Vermaire JH. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health*. 2018;21(10):1234–1242.

9. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HAP, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. *Osteoporos Int*. 2008;19(7):1029–1037.
10. de Bekker-Grob EW, Hol L, Donkers B, et al. Labelled vs unlabelled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value Health*. 2010;13(2):315–323.
11. Kohler RE, Lee CN, Gopal S, Reeve BB, Weiner BJ, Wheeler SB. Developing a discrete choice experiment in Malawi: eliciting preferences for breast cancer early detection services. *Patient Prefer Adherence*. 2015;9:1459–1472.
12. 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.
13. Terris-Prestholt F, Quaife M, Vickerman P. Parameterising user uptake in economic evaluations: the role of discrete choice experiments. *Health Econ*. 2016;25(suppl 1):116–123.
14. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge University Press; 2000.
15. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? *Expert Rev Pharmacoecon Outcomes Res*. 2017;17(6):531–542.
16. Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. *Qual Health Care*. 2001;10(suppl I):i55–i60.
17. Determann D, Korfage IJ, Fagerlin A, et al. Public preferences for vaccination programmes during pandemics caused by pathogens transmitted through respiratory droplets—a discrete choice experiment in four European countries, 2013. *Euro Surveill*. 2016;22(21).
18. Cheraghi-Sohi S, Hole AR, Mead N, et al. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. *Ann Fam Med*. 2008;6(2):107–115.
19. Luce MF, Payne JW, Bettman JR. Emotional trade-off difficulty and choice. *J Mark Res*. 1999;36(2):143.
20. Watson V, Ryan M, Watson E. Valuing experience factors in the provision of chlamydia screening: an application to women attending the family planning clinic. *Value Health*. 2009;12(4):621–623.
21. Adamowicz W, Louviere J, Williams M. Combining revealed and stated preference methods for valuing environmental amenities. *J Environ Econ Manage*. 1994;26(3):271–292.
22. Mueller S, Osidacz P, Francis IL, Lockshin L. Combining discrete choice and informed sensory testing in a two-stage process: can it predict wine market share? *Food Qual Prefer*. 2010;21(7):741–754.
23. Natter M, Feurstein M. Real world performance of choice-based conjoint models. *Eur J Oper Res*. 2002;137(2):448–458.
24. Carlsson F, Martinsson P. Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *J Environ Econ Manage*. 2001;41(2):179–192.
25. Cameron TA, Poe GL, Ethier RG, Schulze WD. Alternative non-market value-elicitation methods: are the underlying preferences the same? *J Environ Econ Manage*. 2002;44(3):391–425.
26. Fifer S, Rose J, Greaves S. Hypothetical bias in stated choice experiments: is it a problem? And if so, how do we deal with it? *Transp Res Part A Policy Pract*. 2014;61:164–177.
27. Mark TL, Swait J. Using stated preference and revealed preference modeling to evaluate prescribing decisions. *Health Econ*. 2004;13(6):563–573.
28. Kesternich I, Heiss F, McFadden D, Winter J. Suit the action to the word, the word to the action: hypothetical choices and real decisions in Medicare Part D. *J Health Econ*. 2013;32(6):1313–1324.
29. Ryan M, Watson V. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health Econ*. 2009;18(4):389–401.
30. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*. 2013;31(4):345–355.
31. Lambooij MS, Harmsen IA, Veldwijk J, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC Med Res Methodol*. 2015;15(1).
32. Salampessy BH, Veldwijk J, Schuit JA, et al. The predictive value of discrete choice experiments in public health: an exploratory application. *Patient*. 2015;8(6):521–529.
33. Lancsar E, Swait J. Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics*. 2014;32(10):951–965.
34. Goodwin K, Viboud C, Simonsen L. Antibody response to influenza vaccination in the elderly: a quantitative review. *Vaccine*. 2006;24(8):1159–1169.
35. Determann D, Korfage IJ, Lambooij MS, et al. Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. *PLoS One*. 2014;9(7):e102505.
36. Burns VE, Ring C, Carroll D. Factors influencing influenza vaccination uptake in an elderly, community-based sample. *Vaccine*. 2005;23(27):3604–3608.
37. Shono A, Kondo M. Parents' preferences for seasonal influenza vaccine for their children in Japan. *Vaccine*. 2014;32(39):5071–5076.
38. Sadique MZ, Devlin N, Edmunds WJ, Parkin D. The effect of perceived risks on the demand for vaccination: results from a discrete choice experiment. *PLoS One*. 2013;8(2):e54149.
39. de Bekker-Grob EW, Veldwijk J, Jonker M, et al. The impact of vaccination and patient characteristics on influenza vaccination uptake of elderly people: a discrete choice experiment. *Vaccine*. 2018;36(11):1467–1476.
40. Sándor Z, Wedel M. Heterogeneous conjoint choice designs. *J Mark Res*. 2005;42(2):210–218.
41. Pachur T, Spaar M. Domain-specific preferences for intuition and deliberation in decision making. *J Appl Res Mem Cogn*. 2015;4(3):303–311.
42. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. *Diabetes Care*. 2008;31(5):874–879.
43. van der Vaart R, Drossaert CHC, Taal E, et al. Validation of the Dutch functional, communicative and critical health literacy scales. *Patient Educ Couns*. 2012;89(1):82–88.
44. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the subjective numeracy scale. *Med Decis Making*. 2007;27(5):672–680.
45. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. *Med Decis Making*. 2007;27(5):663–671.
46. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26(8):661–677.
47. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient*. 2015;8(5):373–384.
48. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York: Academic Press; 1974:105–142.
49. Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law*. 2009;4(4):527–546.
50. Mangham LJ, Hanson K, McPake B. How to do (or not to do) … Designing a discrete choice experiment for application in a low-income country. *Health Policy Plan*. 2009;24(2):151–158.
51. Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. *Pharmacoeconomics*. 2017;35(7):697–716.
52. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health*. 2016;19(4):300–315.
53. Bhat CR. A heteroscedastic extreme value model of intercity travel mode choice. *Transp Res Part B Methodol*. 1995;29(6):471–483.
54. Salisbury LC, Feinberg FM. Alleviating the constant stochastic variance assumption in decision research: theory, measurement, and experimental test. *Mark Sci*. 2010;29(1):1–17.
55. Brownstone D, Train K. Forecasting new product penetration with flexible substitution patterns. *J Econom*. 1998;89(1-2):109–129.
56. Scarpa R, Ferrini S, Willis K. Performance of error component models for status-quo effects in choice experiments. In: Scarpa R, Alberini A, eds. *Applications of Simulation Methods in Environmental and Resource Economics: The Economics of Non-Market Goods and Resources*. Vol. 6. Dordrecht, The Netherlands: Springer; 2005.
57. Bierlaire M. *Estimating choice models with latent variables with PythonBiogeme. Technical Report TRANSP-OR 160628*. Transport and Mobility Laboratory, ENAC, EPFL; 2016.
58. Bierlaire M, Fetiarison M. Estimation of discrete choice models: extending BIOGEME. Paper presented at: Swiss Transport Research Conference (STRC), Ascona, Switzerland. 2009.
59. Train K. *Discrete Choice Methods With Simulation*. Cambridge, MA: Cambridge University Press; 2003.
60. Revelt D, Train K. Customer-specific taste parameters and mixed logit: households' choice of electricity supplier. Working paper. https://cloudfront.escholarship.org/dist/%20prd/content/qt1900p96t/qt1900p96t.pdf. Accessed April 5, 2019.
61. Hess S. Conditional parameter estimates from mixed logit models: distributional assumptions and a free software tool. *J Choice Model*. 2010;3(2):134–152.
62. Hensher D, Rose J, Greene W. *Applied Choice Analysis*. 2nd ed. Cambridge, UK: Cambridge University Press; 2015.
63. Mohammadi T, Bansback N, Marra F, et al. Testing the external validity of a discrete choice experiment method: an application to latent tuberculosis infection treatment. *Value Health*. 2017;20(7):969–975.
64. Wright SJ, Vass CM, Sim G, Burton M, Fiebig DG, Payne K. Accounting for scale heterogeneity in healthcare-related discrete choice experiments when comparing stated preferences: a systematic review. *Patient*. 2018;11(5):475–488.
65. Rolfe J, Bennett J. The impact of offering two versus three alternatives in choice modelling experiments. *Ecol Econ*. 2009;68(4):1140–1148.