Patient (2015) 8:521–529
DOI 10.1007/s40271-015-0115-2

ORIGINAL RESEARCH ARTICLE

The Predictive Value of Discrete Choice Experiments in Public Health: An Exploratory Application

Benjamin H. Salampessy · Jorien Veldwijk · A. Jantine Schuit · Karolien van den Brekel-Dijkstra · Rabin E. J. Neslo · G. Ardine de Wit · Mattijs S. Lambooij

Published online: 25 January 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract
Objective The objective of this study was to assess the predictive value of a discrete choice experiment (DCE) in public health by comparing stated preferences to actual behavior.
Methods 780 Type 2 diabetes mellitus (T2DM) patients received a questionnaire, containing a DCE with five attributes related to T2DM patients' willingness to participate in a combined lifestyle intervention. Panel mixed-multinomial-logit models were used to estimate the stated preferences based on 206 completed DCE questionnaires. Actual participation status was retrieved for 54 respondents based on patients' medical records and a second questionnaire. Predicted and actual behavior data were compared at population level and at individual level.
Results Based on the estimated utility function, 81.8 % of all answers that individual respondents provided on the choice tasks were predicted correctly. The actual participation rate at the aggregated population level was minimally underestimated (70.1 vs. 75.9 %). Of all individual choices, 74.1 % were predicted correctly with a positive predictive value of 0.80 and a negative predictive value of 0.44.
Conclusion Stated preferences derived from a DCE can adequately predict actual behavior in a public health setting.

Keys Points for Decision Makers

To date, very little is known about the extent to which estimated participation rates based on discrete choice experiment (DCE) results accurately predict behavior.

Both at an aggregated population level and at an individual level, high correspondence rates between predicted and actual participation behavior were found.

Additional studies investigating the predictive value of DCEs by comparing stated preferences and actual behavior are urgently needed.

1 Introduction

Discrete choice experiments (DCEs) had been applied for many years in transportation economics, environmental economics, and marketing before being introduced in health economics [1]. Since their introduction, outcomes from DCEs are increasingly being used to inform and support healthcare policy making [1]. However, critics

B. H. Salampessy · J. Veldwijk (✉) · A. Jantine Schuit · G. Ardine de Wit · M. S. Lambooij
Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands
e-mail: jorien.veldwijk@rivm.nl

J. Veldwijk · R. E. J. Neslo · G. Ardine de Wit
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

A. Jantine Schuit
Department of Health Sciences, EMGO Institute for Health and Care Research, VU University, Amsterdam, The Netherlands

K. van den Brekel-Dijkstra
Leidsche Rijn Julius Health Care Centers, Utrecht, The Netherlands

question whether the outcomes of a DCE are a good proxy for actual behavior of patients and consumers. The hypothetical choices (stated preferences) made by respondents in DCEs only reflect what they presumably would have chosen given the specific set of program characteristics (i.e., attributes). These stated preferences may be different from choices made by the respondents in real-life settings (revealed preferences) [2, 3].

Outside of health economics, stated preferences have been compared to actual behavior to some extent already [4–8]. These studies concluded that, in general, stated preferences could predict actual behavior in various study conditions, e.g., different elicitation methods, market segments, and time periods [4–8]. Within health economics, systematic reviews [1, 2, 9] identified only two studies that concerned the predictive value of DCEs [3, 10], and another study was recently published [11]. The first study conducted by Mark and Swait [10], who used both stated and actual behavior in their data analyses, concluded that stated preferences could be used to model market shares of newly introduced medication. However, no direct comparison was made between predicted and actual behavior at the individual level [2]. The one study that compared predicted behavior to actual behavior in a healthcare setting was conducted by Ryan and Watson [3]. They asked women who visited a fertility clinic to indicate whether they were willing to participate in a Chlamydia screening test using multiple hypothetical scenarios. Subsequently, the women were asked to participate in an existing screening test offered at the clinic (this test was identical to one of the scenarios in the questionnaire). In the real-life setting, 81 % of the women behaved in accordance with what they had stated in the questionnaire. Most incorrect predictions concerned women who had stated they would participate in the screening test but then declined this test when it was offered to them. The authors concluded their predictions overestimated actual behavior but emphasized that more research is needed to support their findings [3].

In the literature, several explanations have been suggested for the discrepancies found in most studies that compared stated and actual behavior. Choices in DCEs may not have the same consequences (e.g., in terms of clinical effects, financial costs) for respondents as real-life decisions, which is also referred to as hypothetical bias [12–14]. In addition, it is generally known that, in most cases, individual behavior is not solely based on whether the preferred program characteristics are present. Other circumstantial factors may also affect the decisions of respondents in real-life settings, e.g., suffering from illnesses, lack of time, and lack of local facilities [15, 16]. Although the most important attributes and levels concerning a specific decision are preferably included in DCEs, these studies are limited in the number of attributes by nature; hence, it may not always be possible to include all important attributes [5]. Another explanation is the intention–behavior gap. This is known to cause differences between planned and actual behavior, i.e., some respondents may change their initial intention prior to behavioral execution [17]. Since the outcomes of DCEs are based on stated behavior, the intention–behavior gap may cause incorrect predictions. All of these factors will influence individuals' behavior while not being accounted for in DCE studies. In the DCE literature, this is referred to as scale difference [5, 18]. Due to the fixed-choice contexts and detailed information on a limited number of attributes, there is less 'noise' in stated preference data than in revealed preference data (i.e., in stated preference data the error variance will be lower and the scale will be higher). It can be questioned to what extent stated preferences can accurately predict actual behavior if such scale effects are not or cannot be accounted for.

Since outcomes of DCEs are increasingly used to support public health policy making [2], the extent to which stated preferences predict actual behavior is of societal interest. Therefore, this study aimed to explore the predictive value of a DCE by assessing the consistency between stated preferences retrieved by a DCE and actual behavior in a specific healthcare context, i.e., type 2 diabetes mellitus (T2DM) patients and their participation in a combined lifestyle intervention (CLI).

2 Methods

The study consisted of two stages: (1) stated preferences were derived from a DCE; and (2) actual choices made by respondents in real-life settings were determined. To investigate the predictive value of this DCE, we compared respondents' actual choices about participation with predictions about their participation based on the stated preferences.

2.1 First Stage: Estimating Stated Preferences

All T2DM patients (except those who were terminally ill and those with a mental illness) in four general practices of health centers located in the area of Utrecht, The Netherlands were eligible to participate in this study. A questionnaire (questionnaire A) was sent to these 780 eligible patients by postal mail. Questionnaires were completed on a voluntary basis. A reminder was sent to patients who had not returned the questionnaire after 3 weeks. Questionnaire A contained questions concerning respondents' demographics and health status and ended with the DCE.

2.1.1 Attributes and Levels of the Discrete Choice Experiment (DCE)

The attributes and the levels that were used in the DCE were selected based on literature review, expert interviews, and focus group interviews. These expert and focus group interviews were conducted to (1) determine the most important attributes, and (2) ensure that the attribute levels were considered realistic and consistent with current practice. A detailed description of this process is described elsewhere [19]. The five attributes with the corresponding levels are shown in Table 1. The menu schedule and the physical activity (PA) schedule attributes described the level of guidance provided by a lifestyle coach when establishing the respondents' goals concerning their diets and PA behavior. Respondents set these goals during consultations with their coach. Consultation structure described whether these consultations took place individually or in small or large groups with other T2DM patients. The expected results in terms of weight loss and physical fitness that respondents had before starting the program were reflected in the expected outcome attribute. Finally, the out-of-pocket (OOP) costs attribute reflected the amount that respondents had to pay when they participated in the CLI.

2.1.2 Study Design of the DCE

NGene 1.1 software (ChoiceMetrics Pty Ltd., 2011) was used to create a D-efficient design for this study [18, 20]. The software was instructed to create a design using a panel mixed multinomial-logit model (MIXL), with all beta-priors set at zero, 100 Halton draws and 500 repetitions. It was assumed that there would be no interaction between attributes, while level balance and minimal overlap between attribute levels were optimized. Utility balance between the alternatives within each choice task was optimized to be between 60 and 40 % and 80 and 20 %. The final design (D-error = 0.37) consisted of 18 unique choice tasks divided over two blocks. As shown in the example choice task of the DCE in Fig. 1, each choice task consisted of two unlabeled hypothetical CLI programs and an opt-out option. The latter was included to resemble real-life settings more closely, since patients with T2DM could always decline the offer to participate.

2.1.3 Analysis of DCE Data

Analyses of Eq. (1) were performed using the panel MIXL technique in Nlogit 5.0 (Econometric Software, Inc., Plainview, NY, USA). This technique adjusts for the correlation between the answers within respondents, i.e.,

Table 1 Attributes and corresponding levels included in this discrete choice experiment study[a]

Attributes	Levels
1. Menu schedule	*Flexible (reference)* you set your own goals and develop your own menu schedule to reach these goals without the assistance of a lifestyle coach
	General your lifestyle coach informs you about health and unhealthy foods, using food information and examples of recipes
	Elaborate your lifestyle coach develops a menu schedule that meets your needs and wishes
2. Physical activity schedule	*Flexible (reference)* you set your own goals and develop your own activity schedule to reach these goals without the assistance of a lifestyle coach
	General your lifestyle coach informs you about what physical activities would be good for you, using information about physical activity and examples of exercises
	Elaborate your lifestyle coach develops a physical activity schedule that meets your needs and wishes
3. Consultation structure	*Individual (reference)* the consultations of the lifestyle program are individually
	Consultation 5 the consultations of the lifestyle program are in groups of *5 other patients*
	Consultation 10 the consultations of the lifestyle program are in groups of *10 other patients*
4. Expected outcomes	*No weights loss (reference)* but you feel fitter
	A *weight loss of 5 kg* and you feel fitter
	A *weight loss of 10 kg* and you feel fitter
5. OOP costs[b]	OOP costs of €*75* per 3–6 months
	OOP costs of €*150* per 3–6 months
	OOP costs of €*225* per 3–6 months

OOP out-of-pocket

[a] Attributes and corresponding levels are also described elsewhere [18]

[b] Levels of the linear attribute OOP costs were coded as 0.75 (€75), 1.5 (€150), and 2.25 (€225)

Imagine your General Practitioner or Nurse Practitioner would recommend you to participate in a combined lifestyle intervention program for 3-6 months.
Which situation would you rather prefer to participate, situation 1 or situation 2?
When you rather not prefer to participate in both situations, you can tick the opt-out option

	Lifestyle program 1	**Lifestyle program 2**	Opt-out option
Menu schedule	General	Flexible	None
Physical Activity schedule	General	Flexible	None
Consultation structure	Group of 5 others	Individual	None
Expected outcomes	5 kilograms weight loss and feeling fitter	5 kilograms weight loss and feeling fitter	None
Out-of-the-pocket costs	75 euros	75 euros	0 euros
I would rather prefer to participate in	☐	☐	☐

Fig. 1 Example of a choice task used in the discrete choice experiment

adjusting for the multilevel data structure, as well as preference heterogeneity between respondents [20, 21]. This model is fitted in an iterative manner until the log-likelihood function is optimized.

$$\begin{aligned} U_{nj} = V_{nj} + \varepsilon_{nj} = {} & \beta 0_i + \beta 1 \times \text{General Menu schedule} \\ & + \beta 2 \times \text{Elaborate Menu schedule} \\ & + \beta 3 \times \text{General PA schedule} \\ & + \beta 4 \times \text{Elaborate PA schedule} \\ & + \beta 5 \times \text{Consultation structure in groups with 5 others} \\ & + \beta 6 \times \text{Consultation structure in groups with 10 others} \\ & + \beta 7 \times \text{Expected Outcome weight loss of 5 kilograms}_i \\ & + \beta 8 \times \text{Expected Outcome weight loss of 10 kilograms}_i \\ & + \beta 9 \times \text{OOP-costs} + \varepsilon_{nj} \end{aligned} \quad (1)$$

The latent utility '*U*' of individual '*n*' concerning scenario '*j*' can be estimated by taking the sum of the systematic utility element '*V*' (i.e., the utility of individual '*n*' concerning scenario '*j*' calculated based on all attribute levels and covariates) and the random error term 'ε' (i.e., all unobserved and unobservable factors that influence the utility of person '*n*' concerning scenario '*j*'). This error term follows an extreme value type 1 distribution. $\beta 0$ represented the constant of the model. The constant describes the utility of T2DM patients for a lifestyle program versus no lifestyle program (opt-out) when all attributes are set at zero. $\beta 1$–$\beta 9$ represented the attribute level estimates. Four attributes (menu and PA schedule, consultation structure, and expected outcome) were coded using effects coding. The reference category in effects coding was coded as -1, which summed the attribute in each category to zero. The estimates for the reference categories were calculated using (-1) * (βeffectcode1 + βeffectcode2) [5, 22]. Based on model fit tests (Akaike information criterion, Bayesian information criterion, Log likelihood) it was tested which model fitted best to the data. Based on the significance level of the standard deviation (SD) of the attributes it was tested what attributes should be included as random parameters due to significant preference heterogeneity. In addition, different distributions of the random parameters were tested and, based on the model fit results, all random parameters were included with a normal distribution. The constant variable, expected outcome, and OOP costs were included as random parameters (indicated by *i* in the utility equation) with a normal distribution. Since the panel MIXL model does not account for variability in individual errors (scale heterogeneity) [23], the modeling procedures described above were repeated using a Heteroscedastic Extreme Value (HEV) model (accounts for scale heterogeneity), and a Generalized-Mixed-Logit (G-MIXL) model (accounts for scale and preference heterogeneity) [23].

2.2 Second Stage: Determining Actual Behavior

A CLI that was implemented at the participating health centers was used to determine actual behavior of respondents. This CLI was offered to patients who had cardiovascular disease, T2DM, chronic obstructive pulmonary disease, anxiety or depressive disorders, and to patients who smoked or had body mass index (BMI) scores >30 kg/m^2. While this program was available for patients with a wider range of health conditions, in this study only patients with T2DM were selected. Within the program, goals were set to assist patients to improve their lifestyle with the help of lifestyle coaches, physiotherapists, dieticians, and specialized nurses. Respondents who had completed questionnaire

A ($n = 206$) were eligible for the second stage of the study. After administration of questionnaire A, general practitioner (GPs) or nurse practitioners (NPs) discussed participation in the CLI with these respondents.

2.3 Actual Behavior

In February 2013 actual behavior was determined for all eligible respondents of the second stage using patients' medical records and an additional questionnaire (questionnaire B). Respondents' participation in CLI was defined as having an intake appointment with the lifestyle coach. Respondents were marked as non-participant when they had been offered participation in the program but declined this offer. The one respondent that did make an intake appointment but did not show up was also marked as a non-participant. Fourteen respondents were excluded from this stage of the study due to deregistration from the participating health centers, death, or terminal illness.

2.4 Statistical Analyses

To test the difference in demographic characteristics between the respondents that only completed the first stage of this study and the respondents that completed both the first and second stage of this study, independent sample *t* tests were used. Results were considered statistically significant if $p < 0.05$.

2.4.1 Within-Respondent Consistency

Two within-respondent consistency tests were performed to assess to what extent the stated preferences could reproduce the actual choices made by respondents in the DCE. In both tests, predicted choices were determined using the stated preferences and were then compared to the actual choices of the DCE. In the first test, these stated preferences were based on data from all respondents ($n = 206$). Since only data derived from the questionnaire was used, the test assessed the predictive value of the fitted model itself. To determine which scenario in each choice task respondents would prefer, the individual utility scores that resulted from the MIXL were used. Subsequently, in agreement with the Random Utility Theory [12], utility maximization was assumed in respondents' decision-making process. Therefore, it was expected that the respondent would choose the scenario with the highest utility score within the choice task. The procedure described above was then repeated in a slightly different manner. In the second test, it was tested whether the stated preferences that were measured among a random subgroup (50 %, $n = 103$) of respondents could be used to correctly predict the actual choices of the remainder of the sample for each of the choice tasks.

2.4.2 Predictive Value at Population Level

The predictive value of DCEs at aggregated level was determined by comparing the estimated participation rate based on the stated preferences to the actual participation rate found in the second stage of the study. The participation rate was estimated based on the CLI as actually implemented at the health centers to allow comparison between stated and actual behavior. This CLI consisted of an elaborate menu schedule, a general PA schedule, and an individual consultation structure and was offered free of charge. According to the guidelines of Dutch General Practitioners for managing patients with obesity or T2DM [24], a 5 % weight loss of obese patients provides considerable health gains and is assumed to be realistically achievable without surgical treatment. Due to the current weight status and BMI of the respondents in our sample, a 5 % weight loss equals a weight loss of 5 kg or slightly more. Therefore, the expected outcomes attribute was set at 5 kg. To estimate the participation rate based on the stated preferences, utility scores were calculated. Since random parameters were included, the probability of participation could not be calculated directly. Therefore, the mean probability of 10,000 simulations was estimated by taking the average of all simulated probabilities given every tested CLI scenario, which was calculated as $1/(1 + \exp^{-V})$ [5, 12]. Similar to the within-respondent consistency test, the scenario with the highest utility score was expected to be the choice of the respondent. Finally, outcomes of these analyses were compared to the actual participation rate concerning the CLI that was observed in this study by means of a Chi-square (χ^2) test.

2.4.3 Predictive Value at Individual Level

Finally, the predicted choices of respondents were compared to their actual choices in real-life settings. Individual utility scores were calculated for the CLI that was implemented at the health centers and subsequently compared to the opt-out option. Fixed coefficients of the MIXL were used for three attributes (menu and PA schedule and consultation structure), while individual coefficients were used for one attribute (expected outcome). The fifth attribute (OOP costs) remained at zero. Again, the scenario with the highest utility score was expected to be the choice of respondents. Finally, these predicted choices were compared to the choice respondents made concerning participating in the CLI in the real-life setting. Results were presented as percentages of correctly predicted choices (correspondence level), the proportion of correctly

identified participants [positive predictive value (PPV)], and the proportion of correctly identified non-participants [negative predictive value (NPV)]. In addition, results were also described as in terms of sensitivity and specificity, and the Cohen's Kappa coefficient was calculated.

3 Results

3.1 First Stage: Estimating Stated Preferences

3.1.1 Study Population

The response rate of questionnaire A was 26.4 % ($n = 206$). As shown in Table 2, respondents had a mean age of 63 years (SD 11.4). The majority had attained a medium educational level (43.3 %) and were of West-European origin (64.4 %). On average, respondents had been diagnosed with T2DM almost 8 years prior to completion of questionnaire A (7.8 years, SD 6.2). In addition, respondents had mean glycosylated hemoglobin of 52.6 mmol/mol (6.96 %) (SD 10.1) and a BMI score of 29.7 kg/m^2 (SD 5.4).

Table 2 Total study population ($n = 206$)

	Mean (SD)	%
Patients' demographics		
Age	63.4 (11.47)	
Sex		53.7
Men		
Education level		37.0
Low		
Medium		43.3
High		19.7
Ethnicity		64.4
West-European		
Moroccan, Turkish		8.8
Other or not specified		26.8
Status of disease		
Duration of diabetes mellitus (years)	7.8 (6.22)	
HbA_{1c} (mmol/mol)	52.6[a] (10.07)	
BMI (kg/m^2)[b]	29.7 (5.43)	
Dutch EQ-5D[b]	0.88 (0.18)	

BMI body mass index, *HbA_{1c}* glycosylated hemoglobin, *SD* standard deviation

[a] This equals an HbA_{1c} of 6.96 %

[b] Significant differences ($p < 0.05$) found between the respondents of only the first stage ($n = 152$) and respondents of the first and second stage ($n = 54$) of the study. On average, respondents that participate both in the first and second stage had higher BMI scores [(33.8 kg/m^2 (SD 6.4)] and lower EQ-5D scores [0.81 (SD 0.2)]

3.1.2 Stated Preferences

Of the 1,818 (206 × 9) possible choice tasks, 1,504 were completed by respondents. As shown in Table 3, three attributes showed significant attribute levels estimates (i.e., PA schedule, expected outcome and OOP costs), implying that the two other attributes did not significantly affect the choice for participation in a CLI. The negative coefficient of OOP costs indicates that with a decrease in OOP costs, the willingness to participate in a CLI increases. Respondents preferred an elaborate menu and a general PA schedule over all other menu and PA schedules. Individual consultations were preferred over consultations in groups,

Table 3 Estimates of the attribute levels based on the panel mixed-logit model

Attribute	Estimate	Standard error
Constant		
Mean	−1.09**	0.37
SD	3.32	0.30
Menu schedule		
Flexible (reference)	−0.21	0.12
General	0.06	0.14
Elaborate	0.15	0.16
PA schedule		
Flexible (reference)	−0.16	0.10
General	0.19*	0.08
Elaborate	−0.03	0.10
Consultation structure		
Individual (reference)	0.12	0.13
Groups of 5	0.09	0.12
Groups of 10	−0.21	0.18
Expected outcome		
No weight loss (reference)		
Mean	−0.80**	0.15
SD	1.33	0.18
Weight loss 5 kg		
Mean	0.51**	0.12
SD	0.20	0.19
Weight loss 10 kg		
Mean	0.29	0.19
SD	1.31**	0.19
OOP costs		
Mean	−1.26**	0.23
SD	−1.10**	0.15

The fitted model consisted of four effects-coded attributes and a continuous-coded attribute (i.e., OOP costs). In addition, three parameters (i.e., the constant, expected outcome, and OOP costs) were set random. The SD reflects the variance between the individual coefficients and the average coefficient

OOP out-of-pocket, *PA* physical activity, *SD* standard deviation

* Significant at $p < 0.05$, ** significant at $p < 0.01$

and expected weight loss of 5 kg was preferred over no weight loss or a 10 kg weight loss. Analyses with HEV and G-MIXL models showed no scale heterogeneity since the scale parameters were insignificant in both models (results not shown).

3.2 Second Stage: Determining Actual Behavior

Actual choices were retrieved for 54 respondents based on patients' medical records ($n = 41$) and questionnaire B ($n = 13$). The latter showed a response rate of 43.1 %. All 54 respondents were included in the second stage of study in which 41 respondents (75.9 %) reported that they chose to participate in the CLI, while 13 respondents declined the offer to participate. All demographics of the 54 respondents of the second stage are equal to those of the respondents that only participated in the first stage, except they had higher BMI scores (33.8 kg/m^2, SD 6.4) and lower self-reported health status (0.81 score, SD 0.2) than respondents of the first stage ($n = 152$) (Table 2, footnote b).

3.2.1 Within-Respondent Consistency

Stated preferences in the first test reproduced 81.8 % of the actual answers made by respondents in the DCE. Using the stated preferences of the randomly selected sample (50 % of the total sample) resulted in accurate predictions in 45.0 % of the choice tasks completed by the other half of the population.

3.2.2 Predictive Value at Population Level

When participation rates of the offered CLI were compared at aggregated level, the estimated participation rate based on the stated preferences was somewhat but not significantly lower (70.1 %) than the actual participation rate (75.9 %) ($\chi^2 = 2.45$, $p > 0.05$).

3.2.3 Predictive Value at Individual Level

As shown in Table 4, when stated preferences and actual behavior were compared at individual level, a correspondence level of 74.1 % was found. In addition, the PPV of 0.80 implies that of those respondents who were predicted to participate in the offered CLI, four out of five actually participated in the CLI. Similarly, the NPV of 0.44 implies that 44 % of non-participation was correctly predicted when compared to actual behavior. Most of the incorrect predictions related to respondents who were predicted to participate but declined the offer in a real-life setting. Moreover, the sensitivity was 0.90 and the specificity was 0.35. Despite the number of correct predictions, the Cohen's kappa was insignificant. Since the majority of T2DM patients decided to participate in the CLI (76 %) and only a minority decided not to participate (24 %), the distribution of patients' behavior was highly skewed, which was probably the cause of the kappa coefficient being insignificant [25, 26].

Table 4 Cross table comparing stated with actual behavior at individual level

	Actual choices		Total
	Participation	No participation	
Predicted behavior			
Participation	36	9	45
No participation	5	4	9
Total	41	13	54

Correspondence level, correctly predicted choices = (36 + 4)/54 = 74.1 %; PPV, share of correctly predicted participations = 36/(36 + 9) = 0.80; NPV, share of correctly predicted non-participations = 4/(4 + 5) = 0.44; Sensitivity = 36/(36 + 4) = 0.90; Specificity = 5/(5 + 9) = 0.35; Cohen's kappa = 0.21 (approximate $T = 1.57$, ns); Actual participation rate = 41/54 = 75.9 %

NPV negative predictive value, *ns* not significant, *PPV* positive predictive value

4 Discussion

Comparisons between stated preferences and actual behavior at aggregated population level showed a slight but not statistically significant underestimation for the stated preferences (70.1 vs. 75.9 %). In 74.1 % of the cases, the stated preferences corresponded with actual behavior at individual level, which resulted in a PPV of 0.80 and NPV of 0.44, a sensitivity of 0.90 and a specificity of 0.35.

Although actual behavior could partly be predicted based on the stated preferences elicited by the DCE, a discrepancy was found, namely 25.9 % of the predicted choices differed from actual behavior. Results indicated an overestimation of the stated preference utilities. There were more respondents for whom participation was predicted who actually opted-out than respondents who were predicted to opt-out to but actually participated.

Three distinct reasons may underlie this finding. First, respondents might have incorporated other attributes in their decision concerning participation in a CLI in a real-life setting. Since the actual choice leading to behavior was not limited to the attributes of the DCE but also included all unobserved attributes, predictions that are based solely on the DCE will inevitably lead to some prediction error. Therefore, it is stressed that the attribute (level) selection procedure is deliberate and concise [5, 12, 27]. Although this process was followed closely within this study [19], there is always a possibility that some important attribute was missed, therefore causing a discrepancy between the

calculated participation rates based on the stated preferences and the actual behavior. Future research may focus on possible design- or statistical-related solutions to reduce the error in calculated utilities due to missed attributes. It might, for instance, be explored whether it would be feasible to have individuals decide to add certain attributes from a predefined list to the obligatory attributes within a DCE using an online setting.

Second, respondents' decisions might be affected by different choice contexts. While in the DCE the respondents all evaluated the choice tasks within the same choice context, in real life respondents may differ with respect to, for instance, demographics, psychological determinants (e.g., attitude), and the priorities or skills of their GP and NP to motivate them to participate in the CLI. The presence of a context effect is underlined by the fact that participation rates of CLIs as reported in literature (23–79 %) [28, 29] indicate the overrepresentation of certain groups. This form of selection bias, as was probably present in this study, is most likely always present in practice. For instance, GPs and NPs are often involved in the process of enrolling participants for CLIs, and the degree to which they will motivate and persuade patients to participate might be subjective to their judgment about the extent to which that particular patient might benefit from participation. External factors as described above may influence the relative importance of the attributes within the DCE (e.g., low income might be related to how the OOP costs attribute is valued) and thus may influence the calculated utility for a certain scenario.

Third, the intention–behavior gap probably always accounts for some error in the predictions of DCE. Perceived barriers and facilitators are likely to come into play when individuals actually decided whether to participate in a CLI. These barriers and facilitators might increase or decrease the utility of the opt-out option as compared to the utility of participating in the CLI without changing the relative importance of the attribute levels. For this reason, a complete correspondence level between stated preference and actual behavior may never be possible.

In summary, external factors that are not included in the DCE, but which in real life affect the utility of a particular scenario, cause an unknown discrepancy between the utilities of the stated and revealed preferences. Such differences are known as scale differences. Several initiatives might be undertaken to minimize the influence of scale on stated preferences. First, an online questionnaire that adapts the choice context of the decision to patient-specific characteristics may be used as a tool to mimic the real-life decision setting as closely as possible, and therefore may reduce the gap between the hypothetical and the real-life choice situation of respondents. Second, analytical models that include context-related covariates (e.g., respondent characteristics or context characteristics) might be used. Hybrid models or models that incorporate interaction terms between attribute levels and context factors may theoretically provide more accurate predictions, since these models incorporate the influence of relevant external factors. However, no study is likely to have sufficient power to incorporate all external factors. Sample size of the current study was sufficient for estimating main effects but not for incorporating several interaction terms. Future research that examines the external validity of a DCE should consider conducting hybrid models or accounting for possible interactions when running sample size calculations beforehand.

Results of comparisons in this study at individual level are in line with results of one other healthcare application [3] and the evidence base from other fields of research [4–8]. While the PPV of 0.80 seems promising, the NPV was not better than could have been expected by chance (0.44). However, PPVs and NPVs are affected by the number of respondents that participate or decline to participate [30], and therefore overrepresentation of participating respondents will obviously result in a higher PPV (i.e., the true positives will then always be higher than the false positives). In this study, a relatively large number of respondents decided to participate while only a small number of respondents declined participation. Subsequently, the reported PPV and NPV may be less accurate than anticipated beforehand due to under-sampling of non-participating respondents.

Additionally, most DCEs are used to predict engaging behavior (i.e., the choice for participation) of respondents, e.g., uptake in new preventive programs, while it is less common to predict refraining behavior (i.e., the choice against participation). This implies that DCEs are likely to be more valuable in understanding why people engage in the behavior under study than understanding why people refrain from that behavior. Refraining behavior may be motivated by a combination of other (non-observed) attribute levels and external factors compared with engaging behavior. The PPV and sensitivity of the current study might be considered as good, since actual behavior was correctly predicted in more than three out of four respondents.

A key issue in the application of DCEs in health policy remains how policy makers should use the outcomes of DCEs. When predicting engaging behavior of respondents, stated preferences derived from DCEs can be used to predict actual behavior of respondents. However, when using the outcomes to predict refraining behavior, different research objectives should be formulated, probably different attribute levels should be identified, different external factors should be measured, and different designs should be conducted.

5 Conclusion

In conclusion, stated preferences can adequately predict actual behavior in a public health setting. However, it remains unclear to what extent missed attributes, choice context, and the intention–behavior gap play a part in the discrepancy between stated preference and actual behavior and how these issues can be overcome. Moreover, it is uncertain to what extent DCEs can predict refraining behavior, which is of particular importance when DCE results are translated into policy implications. Future research should assess the predictive value of DCEs in health economics using different approaches (both modeling engaging and refraining behavior) by using different patient groups and different decision contexts. Because refraining behavior is not simply the opposite of engaging behavior, research on this specific topic is called for.

Acknowledgments and author contributions B. H. S. prepared and designed questionnaire B, performed data analyses, and wrote the manuscript. J. V. prepared and designed questionnaire A, conducted expert interviews and focus groups, designed the DCE, and contributed to all stages mentioned above. M. S. L. contributed to all stages mentioned above. A. J. S. critically reviewed the manuscript. R. E. J. N. contributed to the calculations of the analyses and critically reviewed the manuscript. K. B.-D. assisted with data collection and critically reviewed the manuscript. G. A. W was the project leader of this research and was therefore involved in all stages of the first stage of this study. G. A. W. also critically reviewed the manuscript. This research was funded by the National Institute of Public Health and the Environment. None of the authors declared any conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32:883–902.
2. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21:145–72.
3. Ryan M, Watson V. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Econ. 2009;18:389–401.
4. Adamowicz WL (1992) Combining revealed and stated preference methods for valuing environment amenities. Edmonton: Dept. of Rural Economy, Faculty of Agriculture and Forestry, University of Alberta.
5. Louviere JJ, Hensher DA, Swait JD. Stated choice methods: analysis and applications. Cambridge: Cambridge University Press; 2000.
6. Hensher D, Bradley M. Using stated response choice data to enrich revealed preference discrete choice models. Market Lett. 1993;4:139–51.
7. Ben-Akiva M, Morikawa T. Estimation of switching models from revealed preferences and stated intentions. Transport Res Part A Gen. 1990;24:485–95.
8. Adamowicz W, Swait J, Boxall P, et al. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J Environ Econ Manage. 1997;32:65–84.
9. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2003;2:55–64.
10. Mark TL, Swait J. Using stated preference modeling to forecast the effect of medication attributes on prescriptions of alcoholism medications. Value Health. 2003;6:474–82.
11. Krucien N, Gafni A, Pelletier-Fleury N. Empirical testing of the external validity of a discrete choice experiment to determine preferred treatment option: the case of sleep apnea. Health Econ 2014;1. doi:10.1002/hec.3076.
12. Ryan M, Gerard K, Amaya-Amaya M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using discrete choice experiments to value health and health care. Dordrecht: Springer; 2008. p. 13–46.
13. Johnson FR, Mohamed AF, Ozdemir S, et al. How does cost matter in health-care discrete-choice experiments? Health Econ. 2011;20:323–30.
14. Train K, Wilson WW. Estimation on stated-preference experiments constructed from revealed-preference choices. Transport Res Part B Methodol. 2008;42:191–203.
15. Rhodes RE, Plotnikoff RC, Courneya KS. Predicting the physical activity intention–behavior profiles of adopters and maintainers using three social cognition models. Ann Behav Med. 2008;36:244–52.
16. Thomas N, Alder E, Leese GP. Barriers to physical activity in patients with diabetes. Postgrad Med J. 2004;80:287–91.
17. Sheeran P. Intention–behavior relations: a conceptual and empirical review. Eur Rev Soc Psychol. 2002;12:1–36.
18. Hensher DA, Rose JM, Greene WH. Applied choice analysis: a primer. New York: Cambridge University Press; 2005.
19. Veldwijk J, Lambooij MS, van Gils PF, et al. Type 2 diabetes patients' preferences and willingness to pay for lifestyle programs: a discrete choice experiment. BMC Public Health. 2013;13:1099.
20. Bliemer MC, Rose JM. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transport Res Part B Methodol. 2010;44:720–34.
21. McFadden D, Train K. Mixed MNL models for discrete response. J Appl Econ. 2000;15:447–70.
22. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ. 2005;14:1079–83.
23. Fiebig DG, Keane MP, Louviere J, et al. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Market Sci. 2010;29:393–421.
24. Van Binsbergen J, Langens F, Dapper A, et al. NHG-standaard obesitas. Huisarts Wet. 2010;53:609–25.
25. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol. 1990;43:551–8.
26. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43:543–9.
27. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26:661–77.
28. James DV, Johnston LH, Crone D, et al. Factors associated with physical activity referral uptake and participation. J Sports Sci. 2008;26:217–24.
29. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet. 2009;374:1677–86.
30. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994;309:102.