Social Science & Medicine 246 (2020) 112736

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Can healthcare choice be predicted using stated preference data?

E.W. de Bekker-Grob[a,b,*], B. Donkers[a,c], M.C.J. Bliemer[d], J. Veldwijk[a,b], J.D. Swait[a,b]

[a] Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, the Netherlands
[b] Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[c] Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands
[d] Business School, University of Sydney, Sydney, Australia

ARTICLE INFO

Keywords:
Netherlands
Stated choices
Discrete choice experiment
External validity
Prediction
Healthcare choice

ABSTRACT

Lack of evidence about the external validity of Discrete Choice Experiments (DCEs)-sourced preferences inhibits greater use of DCEs in healthcare decision-making. This study examines the external validity of such preferences, unravels its determinants, and provides evidence of whether healthcare choice is predictable. We focused on influenza vaccination and used a six-step approach: i) literature study, ii) expert interviews, iii) focus groups, iv) survey including a DCE, v) field data, and vi) in-depth interviews with respondents who showed discordance between stated choices and actual healthcare utilization. Respondents without missing values in the survey and the actual healthcare utilization (377/499 = 76%) were included in the analyses. Random-utility-maximization and random-regret-minimization models were used to analyze the DCE data, whereas the in-depth interviews combined five scientific theories to explain discordance. When models took into account both scale and preference heterogeneity, real-world choices to opt for influenza vaccination were correctly predicted by DCE at an aggregate level, and 91% of choices were correctly predicted at an individual level. There was 13% (49/377) discordance between stated choices and actual healthcare utilization. In-depth interviews showed that several dimensions played a role in clarifying this discordance: attitude, social support, action of planning, barriers, and intention. Evidence was found that our DCE yields accurate actual healthcare choice predictions if at least scale and preference heterogeneity are taken into account. Analysis of discordant subjects showed that we can even do better. The DCE measures an important part of preferences by focusing on attribute tradeoffs that people make in their decision to participate in a healthcare intervention. Inhibitors may be among these attributes, but it is more likely that inhibitors have to do with exogenous factors like goals, religion, and social norms. Con-ducting upfront work on constraints/inhibitors of the focal behavior, not just what promotes the behavior, might further improve predictive ability.

1. Introduction

Discrete choice experiments (DCEs) have been introduced in health economics to elicit preferences for health and healthcare (Ryan, 2004). The DCE technique is mainstream in marketing, transport, and environmental economics, where it is used to predict individual and collective choices and calculate willingness-to-pay measures (Bliemer and Rose, 2011; Mahieu et al., 2017). Over the last two decades, the DCE application in health economics has grown exponentially as this method for collecting choice data is relatively easy to apply, appears efficient, and the analysis of the data is able to address a wide range of important questions for regulators, industry, health technology assessment bodies, and patient organizations (de Bekker-Grob et al., 2012; Clark et al., 2014; Soekhai et al., 2019; Rowen et al., 2018; de Bekker-Grob et al., 2008, 2010; Kohler et al., 2015). Among other topics, DCEs in health economics are used for valuing and investigating trade-offs between health and non-health outcomes, developing priority setting frameworks, and predicting uptake for where there is no information or trial data (Soekhai et al., 2019). However, the lack of evidence about the external validity of DCEs is one of the barriers that inhibits greater use of DCEs in healthcare decision-making.

The external validity of DCEs is concerned with ensuring the comparability of hypothetical (stated) choices and actual (revealed) choices (Quaife et al., 2018). To support claims based on DCEs for healthcare decision-making, there is a great need for research on the external validity of DCEs, particularly empirical studies assessing predicted and revealed preferences of a representative sample of participants (Quaife et al., 2018). Although a DCE succeeds in demonstrating internal validity (e.g., how accurately preferences are measured; the extent to which the results are consistent with a priori expectations; and the

* Corresponding author. Erasmus School of Health Policy & Management, Erasmus University Rotterdam, P.O. Box 1738, 3000, DR, Rotterdam, the Netherlands.
E-mail address: debekker-grob@eshpm.eur.nl (E.W. de Bekker-Grob).

https://doi.org/10.1016/j.socscimed.2019.112736
Received 22 July 2019; Received in revised form 8 November 2019; Accepted 13 December 2019
Available online 18 December 2019
0277-9536/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

extent to which the DCE takes account of all things deemed important in the construct's domain (Janssen et al., 2017; Cheraghi-Sohi et al., 2008; Determann et al., 2016; Ryan et al., 2001), this does not guarantee external validity (Ryan et al., 2001; Luce et al., 1999; Watson et al., 2009).

The studies, mainly outside healthcare, that have investigated the external validity of DCEs (Adamowicz et al., 1994; Kesternich et al., 2013; Mueller et al., 2010; Mark and Swait, 2004; Cameron et al., 2002; Carlsson and Martinsson, 2001; Fifer et al., 2014; Natter and Feurstein, 2002), all focus on final outcomes only (Lancsar and Swait, 2014). However, the investigation of the external validity should be much broader (Lancsar and Swait, 2014). That is, we need to know where exactly discrepancies arise when stated choices derived from DCE do not match actual healthcare utilization. Identifying the determinants of DCE the external validity is crucial to obtain predictions with the greatest accuracy (Quaife et al., 2018), and as such, to be valuable to healthcare decision-making. The aim of this study is to measure the external validity of predictions based on stated choices, unravel its determinants, and provide some evidence about whether healthcare choice is predictable.

2. Methods

For several reasons we focused on the field of influenza vaccination for persons 60≤ years of age: (a) subjects face a real choice since they are not obliged to opt for vaccination, which is vital to test the consistency between stated and actual healthcare choices; (b) the number of people facing this decision is large, contributing to the relevance and feasibility of the study; and (c) logistic reasons since we had access to actual influenza vaccination choice data on an individual level, which is crucial to detect individuals who show discordance between stated choices and actual healthcare utilization. A mixed methods approach was used, which included six steps: i) a literature study, ii) expert interviews, iii) focus groups, iv) a survey including a DCE, v) field data, and vi) in-depth interviews. A within-sample design was used (i.e., the same respondents were involved in the survey, in the field data as well as in the in-depth interviews of our study) combining techniques from econometrics, medical and behavioral research. Approval for the study was obtained from the Medical Ethics Committee, Erasmus MC (MEC-2016-095).

2.1. Literature study, expert interviews, and focus groups

We used a literature search (Sadique et al., 2013; Determann et al., 2014; Shono and Kondo, 2014; Goodwin et al., 2006; Burns et al., 2005), interviews with experts in the field of influenza vaccination (n = 4) and three focus groups with patients aged 60 years and older from two Dutch general practices from urban and rural areas (n = 21; i.e., the target group) to develop and operationalize influenza vaccination attributes. During this qualitative work, the nominal group technique (Delbecq and Van de Ven, 1971) was applied to detect the most relevant influenza vaccination attributes to be included in the DCE: vaccination effectiveness, risk of severe side effects, risk of mild side effects, protection duration, and incubation time (Table 1) (de Bekker-Grob et al., 2018). The levels for each attribute incorporated the range of possible vaccination outcomes based on current literature and near future or plausible expectations (Table 1). The levels of the effectiveness attribute (i.e., 20%, 40%, 60%, and 80%) were explained as follow: “Suppose that all 100 people had the flu vaccination, the number of people who get the flu depends on the effectiveness of the flu vaccination. The effectiveness differs per flu vaccination and may, for example, have the following level: 80%. This means that of all 100 people who would get the flu, 80 people will NOT get the flu anymore, while 20 people will still get the flu”. Noteworthy, in the Netherlands, recipients of 60 years and older (our target population) do not have to pay for influenza vaccination.

Table 1
Attributes and attribute levels.

Vaccination attributes	Levels
Effectiveness (%)	20 - 40 – 60 - 80
Risk of severe side effects (x out of 1,000,000)	1 - 10 - 100
Risk of mild side effects (x out of 10)	1 - 3 - 5
Protection duration (months)	3 - 6 - 12
Vaccine will become active (x weeks after vaccination)	2 – 4

2.2. Discrete choice experiment (DCE)

The combination of five attributes with two, three and four levels each results in 216 ($2^1 \times 3^3$ x 4^1) potential influenza vaccination profiles, and in 23,220 (i.e., 216 x 215 x 0.5) unique paired comparisons of influenza vaccination scenarios (i.e., choice tasks). To create a much smaller subset of choice tasks with little loss of information or estimation precision, while accommodating substantial respondent hetero-geneity, a Bayesian D-efficient heterogeneous DCE design was used (Sándor and Wedel, 2005). We generated a heterogeneous DCE design consisting of 10 sub-designs using Fortran language. Together these sub-designs were optimal to estimate a multinomial logit model, based on a main-effects utility function including 2-way interactions between the attribute ‘effectiveness’ and the other attributes. The prior preference information (attribute weights) as required for the Bayesian efficient optimization approach was obtained from best guess priors using expert judgement, and updated after a pilot run of 100 respondents. Each respondent was assigned randomly to a sub-design containing 16 discrete choice tasks each.

The DCE design contained three alternatives for each choice task (i.e., two flu vaccine alternatives and one ‘no flu vaccine’) (see Fig. 1 for an example). Although presenting a choice task with two alternatives (i.e., one flu vaccine alternative and one ‘no flu vaccine’ alternative) is a better reflection of the actual decision for influenza vaccination, previous research showed that DCE models seem to be better able to predict choices and performed best when three alternatives instead of two alternatives are presented to respondents (de Bekker-Grob et al., 2019). Further, showing two flu vaccine alternatives ensures that for respondents who always get vaccinated we still capture their trade-offs with respect to the underlying attributes. The ‘opt out’ (i.e., no flu vaccine) alternative was necessary as influenza vaccination is a preventive intervention and, as in real life, respondents are not obliged to get vaccinated against influenza.

2.3. Survey and sample

Besides the 16 DCE choice tasks (for the estimation of the decision model) and questions regarding respondents’ characteristics, an additional choice task was added to the survey mimicking the real-world choice (see Fig. 2); this additional choice task was placed in the middle of the survey (i.e., after the first 8 DCE choice tasks). Adding this stated choice task mimicking a real-world healthcare decision, we are able to check whether DCE is capable of predicting the choices in a hypothetical situation representing the real-life choice, which is a minimal requirement for the external validity (i.e., if DCE fails here, the external validity will fail as well) (Lancsar and Swait, 2014). The information we presented about the attributes and their levels, was consistent with the Dutch national flyer and invitation that participants would receive from their GP to keep information asymmetry between the hypothetical situation representing the actual decision and the actual decision as small as possible. (The ten different versions of the questionnaire are available in Dutch from the authors on request).

Calculation of optimal sample sizes for a DCE is complicated as it depends on the true values of the unknown parameters in the assumed discrete choice model (Lancsar and Louviere, 2008). Using the rule of thumb proposed by Johnson and Orme (Johnson et al., 2003; Orme,

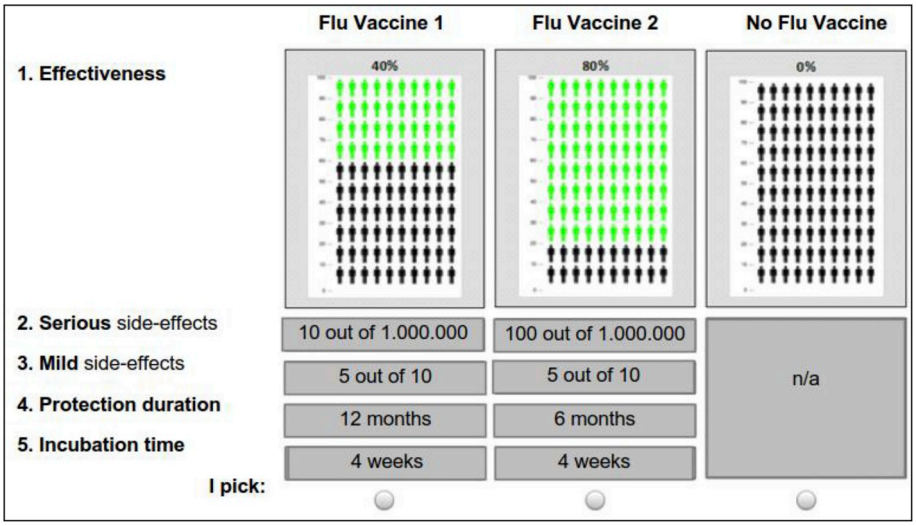

Fig. 1. DCE choice task.

1998), followed by the sample size calculation approach of De Bekker-Grob et al. (de Bekker-Grob et al., 2015) after the pilot run of 100 respondents, a minimal sample size of complete data from 300 respondents was required. Assuming a response rate of 30% (Watson et al., 2017), taking into account a survey completion failure rate of 20%, at least 1250 individuals aged 60 years and older, representative in terms of age and gender, should be recruited. However, given the research budget at hand, we strived to simply maximize the sample size and therefore contacted 1600 individuals aged 60 years and older, representative in terms of age and gender, for the survey via two Dutch general practices mentioned earlier. Trying to overpower the study as much as possible is beneficial for reasons other than statistical precision (e.g., to facilitate in-depth analysis) (de Bekker-Grob et al., 2015).

As the aim was to investigate whether stated vaccination choice can predict actual vaccination choice, each invitee received the survey three weeks before the vaccination invitation letter. This time frame was chosen to give individuals sufficient time to fill out the questionnaire, while the impact of the questionnaire on preferences for the actual choice is limited.

2.4. Field data

Three months after the vaccination invitation letter, field data (i.e., *actual* influenza vaccination data) was collected at an individual level via the same two Dutch general practices as mentioned above. Field data among the same invitees who filled out the questionnaire were collected, which was the only way to test to what extent actual vaccination is consistent with predictions based on stated choices at an individual level. Noteworthy, in the Netherlands, the general practitioner sends the invitation letter for influenza vaccination. There are, however, employers who offer influenza vaccination for their employees as well. In case a person receives two offers (i.e., one from his/her general practitioner and one from his/her employer) and decides to receive vaccination at his/her work, common practice is that this person will

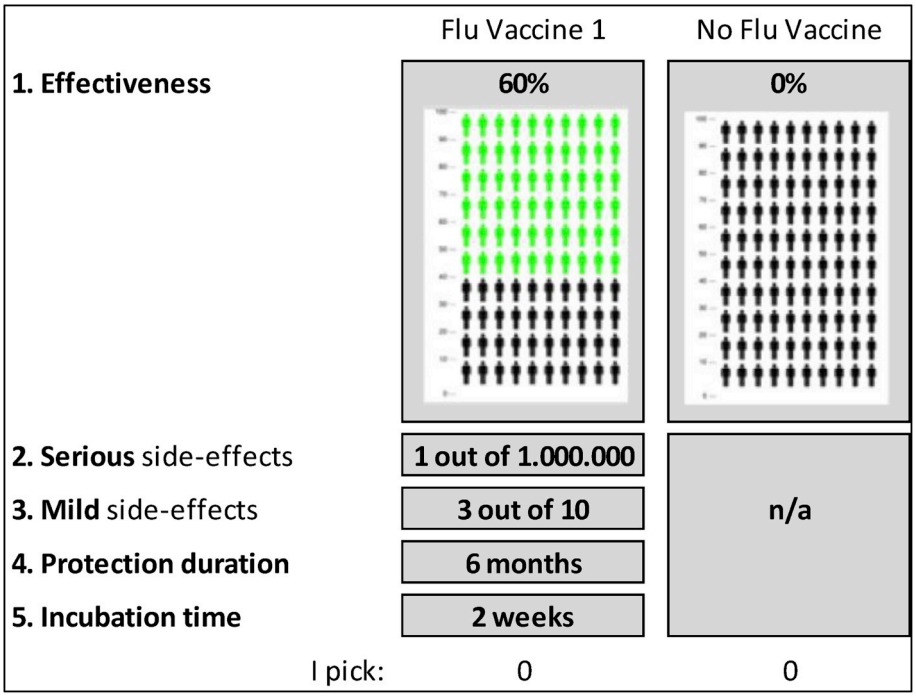

Fig. 2. Choice task mimicking the real-world choice of influenza vaccination.

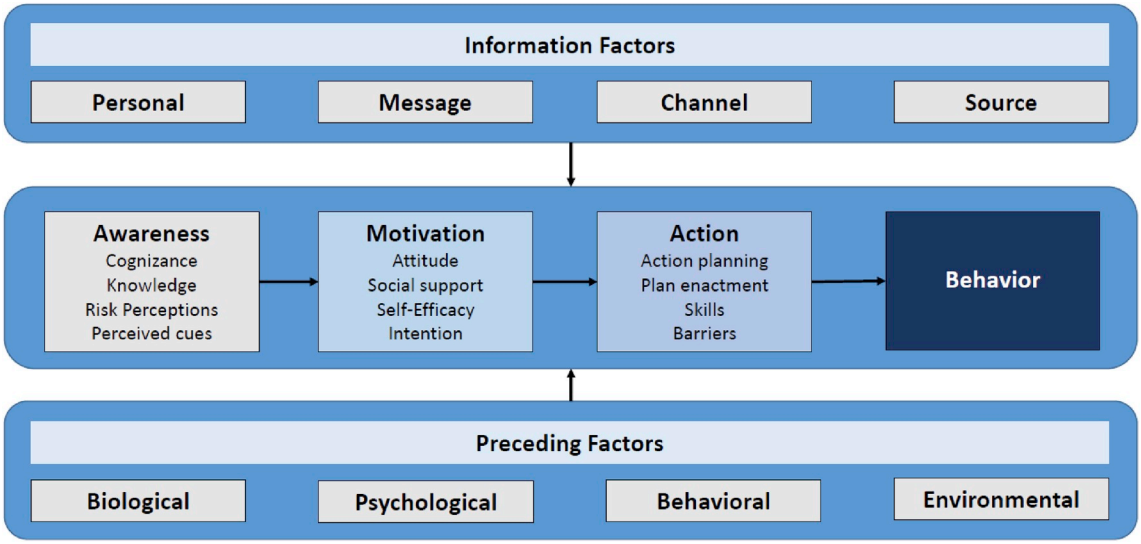

Fig. 3. I-change model (de Vries et al., 2005; de Vries, 2017).

inform his/her general practitioner. That is, this person will fill out the standard influenza vaccination reply card from the general practitioner (i.e., tick the box 'I will receive/received my flu vaccination via my employer') and send the card back to his/her general practitioner. As a result, the electronic record from the general practice will show that this person has been vaccinated.

2.5. In-depth interviews

In-depth semi-structured telephone interviews, based on the Integrated-change (I-change) model (Fig. 3), (de Vries et al., 2005; de Vries, 2017) took place with individuals who showed discordance between stated choices and actual healthcare utilization. The I-change model combines five scientific theories regarding choice behavior change (trans-theoretical model (Prochaska and Diclemente, 1986), the theory of planned behavior (Ajzen, 1991), goal setting theory (Locke and Latham, 2002, 2006), health belief model (Janz and Becker, 1984), and social cognitive theory (Bandura, 1986)). When the team used these theories, insights were obtained about which dimension or dimensions (e.g., risk perception, illness of family/friends, GP's support in the decision-process) could explain the discordance.

2.6. Analysis

To avoid that any missing information (e.g., incomplete DCE data for one or more respondents or missing respondents' characteristics) would have biased the prediction results, respondents without missing values in the survey and vaccination registration were included in the DCE analyses only. An important starting point in investigating the external validity of DCEs is to focus on the role of the researcher who must decide on the model specification. As we focus on analyzing DCE data and their ability to predict actual healthcare choices, we should keep in mind that respondents in stated choice studies probably have a different amount of information about the healthcare interventions they are evaluating than would likely be the case in revealed preference data. Therefore to keep the choice context constant, we first determined 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; see Step A, Fig. 4) needs to be to predict the 'single choice task mimicking the real-world choice' of influenza vaccination (see Figs. 2 and 4) via DCE correctly at an aggregate and individual level. That is, we analyzed the DCE data in a systematic sequence: from a simpler random-utility-maximization (RUM) model (Model A; Table 2) to more and more complex RUM models (Models B-D; Table 2), and from a simpler random-regret-minimization (RRM) model (Model E; Table 2) to more and more complex RRM models (Models F-H; Table 2).

Based on common practice in health economics (de Bekker-Grob et al., 2012; Clark et al., 2014; Soekhai et al., 2019), we started with the homogenous preference, homoscedastic multinomial logit (MNL; Model A, Table 2) (McFadden, 1974) using RUM as the decision rule:

$$U_{in} = V(X_{in},\beta) + \varepsilon_{in}, i \in C \qquad \text{[Eq. 1]}$$

As shown in Equation (1), the latent utility of an alternative i in a choice set C (as presented to individual n) is decomposable into two additively separable parts: (i) a systematic (explainable) component specified as a function of the attributes of the alternatives $V(X_{in}, \beta)$; and (ii) a random (unexplainable) component ε_{in} representing stochastic variation in choices. The MNL model has three key properties: (i) error terms are assumed independently and identically distributed and Extreme Value type I across alternatives and decision makers; (ii) independence of irrelevant alternatives (IIA), following directly from (i); and (iii) no un-attributable 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 (heteroscedastic multinomial logit (HMNL; Model B, Table 2)) followed by relaxing preference homogeneity (with 17 known (reported) subject characteristics (HMNL+; Model C, Table 2)) and by taking randomness of the alternative specific constant parameter to account for stochastic preferences (i.e., heterogeneity from unknown sources) toward the opting out alternative (HMNL++; Model D, Table 2). That is, in Model D we aimed to explain heterogeneity by including subject characteristics *and* using choice responses to obtain individual conditional parameters to explain heterogeneity regarding the alternative specific constant parameter (i.e., $\hat{\alpha}_n$). In this way, we are able to distinguish systematic heterogeneity from stochastic heterogeneity. Noteworthy, we adopted the terminology from Hess & Train where scale heterogeneity and preference heterogeneity are both sources of correlation that reflect random heterogeneity (Hess and Train, 2017). Hence, the following random-utility-maximization choice processes with scale and/or preference heterogeneity and/or random opt-out utility were used (see equations (2)–(5)):

$$U(opt-out)_{in} = \alpha_n + \varepsilon_{in}$$
$$U(opt-in)_{in} = \beta X_i + \delta Z_{1n} X_i + \varepsilon_{in} \qquad \text{[Eq. 2]}$$

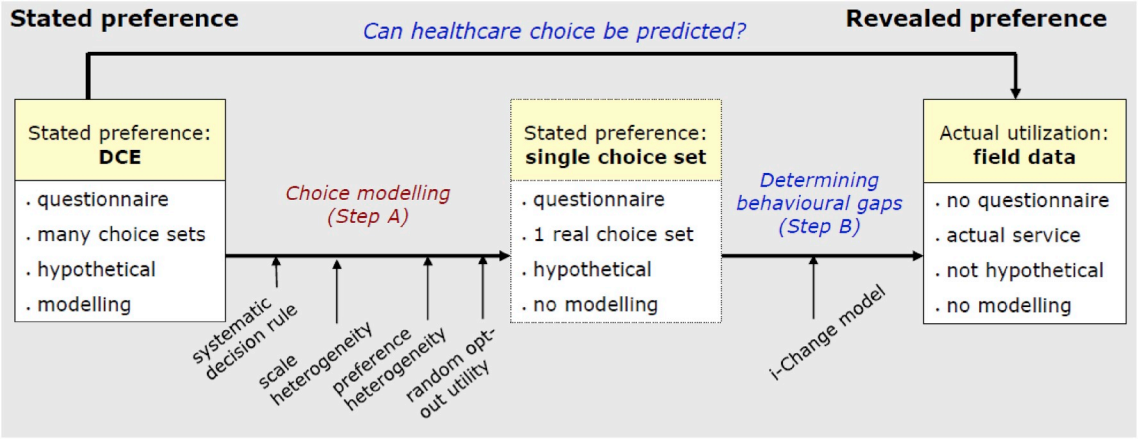

Fig. 4. Comparing stated preferences with actual healthcare utilization.

Table 2
Systematic choice modelling approach.

	Model	A	B	C	D	E	F	G	H
1	Decision rule								
	Random Utility Maximization	X	X	X	X				
	Random Regret Minimization					X	X	X	X
2	Systematic scale heterogeneity		X	X	X		X	X	X
3	Systematic preference heterogeneity			X	X			X	X
4	Random opt-out utility				X				X

where

$$\alpha_n = \overline{\alpha} + \theta Z_{1n} + \eta_n$$
$$\eta_n \sim N(0,\sigma^2) \quad \text{[Eq. 3]}$$

$$\varepsilon_{in} = \text{HEV}(\mu_n) \quad \text{[Eq. 4]}$$

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

The quantity α_n is the alternative specific constant for opting out of vaccination compared to opting into vaccination, Z_1 and Z_2 are two sets of subject characteristics, η_n a normally distributed random component in α_n, and HEV is a Heteroscedastic Extreme Value distribution with scale parameter μ_n (Bhat, 1995; Salisbury and Feinberg, 2010). To employ terminology common to discrete choice models, our full specification is a heteroskedastic error component model (Scarpa et al., 2005; Brownstone and Train, 1998).

The exact same procedure (see Table 5, models E-H) was followed for using RRM as decision rule:

$$R_{in} = \frac{1}{\xi} \sum_{\substack{k \neq i \\ k \in C}} \sum_{a=1}^{A} \log(1 + \exp(\xi \beta_\alpha^* (x_{k,\alpha} - x_{i,\alpha}))) + \varepsilon_{in}, \quad i \in C \quad \text{[Eq. 6]}$$

As shown in Equation (6), the latent regret of an alternative i in a choice set C (as presented to individual n) is the systematic (explainable) component specified as a function of the attributes of the alternatives and a random component ε_{in}. The term $\log(1 + \exp(\xi \beta_\alpha^* (x_{k,\alpha} - x_{i,\alpha})))$ is the core of equation (6) and it quantifies the regret that is associated with comparing the focal alternative i with another alternative k in terms of attribute a. Preference heterogeneity through observed variables Z_{1n} is captured by defining $\beta_\alpha^* = \beta_\alpha + \delta_\alpha Z_{1n}$, similar to equation (2). The chosen functional form implies that regret approaches zero when alternative k performs (much) worse than i in terms of attribute a and that it grows as an approximately linear function of attribute importance β_α^* and the difference in performance between alternatives on the attribute, in case al-ternative i performs (much) worse than k. This specification of regret follows the formulation of Bliemer et al., 2014, 2017, where ξ represents the 'hardness' of the maximum operator in the definition of regret (set to 1 in most previous work). Note that Van Cranenburgh et al. (van Cranenburgh et al., 2015) independently derived the same model using a parameter that represents $1/\xi$. PythonBiogeme software was used to estimate the models (Bierlaire, 2016). The estimable pa-rameter β_α and δ_α represent the main and interaction effects for (the approximation of) the slope of the regret-function for attribute x_α. Like the RUM approach, the latent regret of an alternative also has a random (unexplainable) component ε_{in}. This term represents the stochastic portion of regret and is assumed to be independently and identically distributed (IID) and Extreme Value Type I across alternatives.

To determine the DCE's ability to predict actual healthcare choices at an aggregate and individual level, we first determined for the 'single choice task mimicking the real-world choice' which proportion of the sample opted for influenza vaccination (i.e., the observed hypothetical influenza vaccination uptake); the same procedure was followed for the actual healthcare utilization (i.e., the observed actual influenza vaccination uptake). Second, we determined to what extent the predicted uptake was in agreement with the observed uptake at an aggregate and at an individual level using probability rules, mean values, and 95% confidence intervals, while at an individual level for Models D and H conditional parameter estimates were taken into account as well. Regarding the latter, using the coefficients of Model D, the software PythonBiogeme, the conditional parameter approach of Train (2003) and Revelt & Train (Revelt and Train, 2000), and Excel (Hess (2010)), we were able to determine the alternative specific constants ($\hat{\alpha}_n$) per individual. These alternative specific constants were added to the dataset as a variable, because if individuals have system-atically different preferences, which are unrelated to observed characteristics, not addressing this heterogeneity will bias the estimates of the preference weights. The utility function of Model D (or Model H), incorporating the alternative specific constants per individual, was used to determine the utility (or regret) weight for each individual for influenza vaccination in the 'single choice task mimicking the real-world decision'. Third, we simulated the probabilities for each individual that s/he would opt-in for influenza vaccination. Fourth, we summarized the probabilities for influenza vaccination for respondents who opted for vaccination in the choice task mimicking the real world. Fifth, 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 on an individual level (Hensher et al., 2015). The rationale for employing probability rules in predicting choices here is crucial from an econometric point of view (Hensher et al., 2015). Although in some disciplines, such as marketing, assigning a probability of 1.0 to the highest probability and zero otherwise are

often seen, this would strictly violate the whole idea of a probabilistic choice model associated with random utility maximization (or random regret minimization).

The in-depth semi-structured telephone interviews were recorded and the data was captured by marking respondents' answers to questions that aimed to measure the specific elements of one or more dimensions of the I-Change Model (Step B, Fig. 4) on a pre-specified answer sheet. A relatively high or low presence (or absence) of a dimension could indicate the behavioral gaps between stated preferences and revealed preferences, and as a result the misspecification of the DCE that restricts the predictive ability of DCE for actual healthcare utilization.

3. Results

3.1. Respondents

Of the 1600 invited subjects, 499 (31%) filled out the questionnaire (see Fig. 5). Seventy-six percent (377/499) of the respondents did not have any missing values in the survey nor in the actual healthcare utilization data, and were of interest for this study. These respondents had a mean age of 70 years (SD = 6.6), about 52% were male, and 27% had a higher educational level (Table 3). Approximately 80% of the respondents reported that they were in good health, 14% had experienced influenza (symptoms) last year, and 54% of the respondents mentioned that they had been vaccinated against influenza last year. The number of respondents with strict preferences against vaccinations (i.e., always chose the opt-out alternative in all 16 DCE tasks) was 91/377 (24.1%). The number of respondents with strict preferences for vaccinations (i.e., never chose the opt-out alternative in any of the 16 DCE tasks) was 169/377 (44.8%). The discordance between stated choices and actual healthcare utilization - based on the comparison of the response on the single choice task mimicking the real-world choice vs. actual receipt of the vaccine was 13% (49/377). There were no significant differences between respondents who showed or did not show discordance, except that respondents who showed discordance were more impacted by family ($p = 0.05$) and by certain convictions ($p = 0.08$) regarding influenza vaccination decision (Table 3).

3.2. DCE results

For all four RUM models, the directions of the significant coefficients of the vaccination attributes and its 2-way interactions were consistent with our a priori hypotheses, which implies the theoretical va-lidity of the models (Table 4). 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 RUM model with the best model fit (Model D; i.e., heteroscedastic in error components model), there is evidence of scale and preference heterogeneity. Re-garding scale heterogeneity, respondents who did not have a chronic disease were more consistent in their choices (i.e., a positive parameter estimate indicates greater choice consistency), i.e., they were less affected by random variation in utility than respondents who had a chronic disease ($p < 0.01$). The same result was found for respondents who had a more deliberative decision-making style ($p = 0.06$). Respondents who were impacted by family members seemed to be less consistent in their choices than respondents who were not impacted by family members in their choice for influenza vaccination ($p = 0.06$). Regarding preference heterogeneity, 14 respondent characteristics were observed that explained the preference heterogeneity found ($p < 0.05$): age, gender, education, household, health literacy, numeracy, decision-making style, health state, having a chronic disease, general attitude to vaccination, having been vaccinated last year, experience with side effects of vaccination, the level of impact by family members, and having visited a hospital for him/herself in the last month. For RRM models, the attribute coefficients had the expected signs. The best model fit was also found for the heteroscedastic in error components model (Model H) (Table 5). This model, like Model D, also showed evidence that respondents who did not have a chronic disease were more consistent in their choices ($p < 0.01$). Respondents who were impacted by family members or were higher educated were less consistent in their choices ($p = 0.09$ and $p = 0.05$ respectively). Regarding preference heterogeneity, the same respondent characteristics as Model D were detected that explained the preference hetero-geneity found ($p < 0.05$), except for 'having visited a GP for him/herself in the last month'. There was no evidence that the family members, age or hospital visit in the last month explained the preference heterogeneity found in the RRM model.

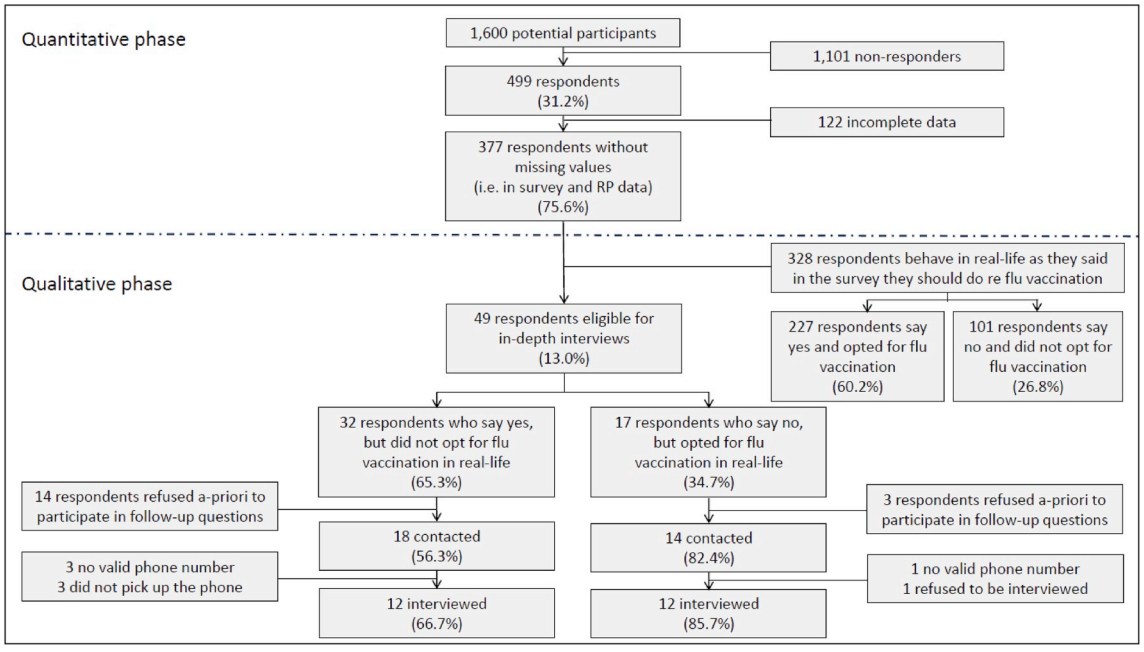

Fig. 5. Flow chart of including participants for the quantitative and qualitative phase.

Table 3
Respondents' characteristics of influenza vaccination survey.

	All respondents		Respondents without discordance between SP and RP		Respondents with discordance between SP and RP		Difference[a]
	N = 377	(%)	N = 328	(%)	N = 49	(%)	p-value
Male	197	52.2	170	51.8	27	55.1	0.67
Age (mean; sd)	70.5	6.6	70.5	6.4	70.7	7.6	0.84
Aged 60–64 years	78	20.7	69	21.0	9	18.4	0.85
Aged 65 years or older	299	79.3	259	79.0	40	81.6	
Education							0.42
Low	177	47.0	158	48.2	19	38.8	
Medium	97	25.7	82	25.0	15	30.6	
High	102	27.1	87	26.5	15	30.6	
Household (living alone; yes)	73	19.4	68	20.7	5	10.2	0.12
Health							0.66
Good	300	79.6	259	79.0	41	83.7	
Moderate	70	18.6	63	19.2	7	14.3	
Bad	7	1.9	6	1.8	1	2.0	
Visited GP last month (yes)	105	27.9	91	27.7	14	28.6	0.87
Visited Hospital last month (yes)	97	25.7	80	24.4	17	34.7	0.16
Suffering from any disease (no)	219	58.1	191	58.2	28	57.1	0.88
Influenza (symptoms) last year (yes)	53	14.1	44	13.4	9	18.4	0.38
Vaccinated against influenza last year (yes)	204	54.1	179	54.6	25	51.0	0.65
Impact certain conviction on flu vaccination (yes)	40	10.6	31	9.5	9	18.4	0.08 *
Vaccination experience effectiveness (bad)	68	18.0	57.0	17.4	11	22.4	0.41
Vaccination experience side effects							0.75
None	208	55.2	181	55.2	27	55.1	
Mild	63	16.7	60	18.3	8	16.3	
Severe	24	6.4	19	5.8	4	8.2	
Family impacts influenza decision (yes)	22	5.8	16	4.9	6	12.2	0.05 **
General attitude vaccination							0.16
In favour	167	44.3	151	46.0	16	32.7	
Neutral	122	32.4	100	30.5	22	44.9	
Against	88	23.3	77	23.5	11	22.4	
Health literacy							
Average (mean; sd)	2.7	0.4	2.7	0.4	2.6	0.5	0.98
Good health literacy (scored 3 or higher)	108	28.6	97	29.6	11	22.4	0.31
Numeracy							
SNS average (mean; sd)	4.2	1.1	4.2	1.1	4.0	1.3	0.30
Objective scores correct (yes)	295	78.2	257	78.4	38	77.6	0.86
Good numeracy (i.e. SNS ≥4 and obj scores correct)	210	55.7	184	56.1	26	53.1	0.75
Decision style							
Decision style average (mean; sd)	2.8	0.7	2.8	0.7	2.8	0.6	0.73
Rather deliberative (3 <)	75	19.9	67	20.4	8	16.3	0.68
neutral (3)	85	22.5	72.0	22.0	13	26.5	
Rather intuitive (< 3)	217	57.6	189.0	57.6	28	57.1	

*p < 0.10; **p= < 0.05.
[a] Difference between respondents' characteristics with and without discordance between stated preference (SP) and revealed preference (RP).

3.3. Predictive ability of DCE

At an aggregate level the SP choice to opt for influenza vaccination was correctly predicted (i.e., taking the 95% confidence interval into account) by DCE models if a RUM model with scale and observed preference heterogeneity was used (Model C; Table 6) or an RRM model with scale, preference heterogeneity and a random opt-out utility was used (Model H; Table 6). For both decision processes (RUM and RRM), we found that a better model fit did not automatically mean better prediction. For instance, Model C led to a better prediction at an aggregate level than Model D, although the in-sample model fit of Model D was significantly better.

At an individual level, the choice for influenza vaccination was predicted best using a heteroskedastic error component model that took into account observed preference heterogeneity through subject characteristics and unknown subject characteristics sources that systematically affect the preference for opting out: 90.8% and 91.1% of the individuals' choices were correctly predicted using a RUM and RRM decision process respectively (i.e., Model D and Model H). Among all models, RRM Model H resulted in the highest positive predictive value (PPV) of 0.96, and RUM Model D resulted in the highest negative predictive value (NPV) of 0.86.

3.4. Discordance between stated choices and actual healthcare utilization

Thirteen percent (49/377) of the respondents showed discordance, of which the majority (32 out of 49; 65%) said that they would opt for influenza vaccination, although they did not opt for influenza vaccination in real-life (Fig. 5). There were no significant differences between respondents who said 'yes' but did not opt for flu vaccination in real-life and respondents who said 'no' but opted for flu vaccination in real-life, except that respondents who said 'yes' but did not opt for flu vac-cination in real-life had a better health condition (p = 0.04).

In-depth interviews based on 24 interviewees showed that the discordance was incorrectly labelled for five out of 24 respondents (21%) (Table 7). That is, i) four respondents made an appointment in the winter instead of the autumn season since they lost their influenza vaccination invitation; as a result, they received an influenza vaccination outside the actual data collection period of this study; and ii) one respondent confirmed that she did not receive an influenza vaccination, although the field data (GP registration) showed that she had received an influenza vaccination. Focusing on the 19 interviewees who confirmed their discordance, we found that several dimensions played a role: attitude, social support, action of planning, barriers, and intention. That is, most subjects who said that they would opt for influenza

Table 4
DCE RUM-results influenza vaccination survey.

	Model A		Model B		Model C		Model D	
	MNL model		HMNL model		HMNL model + systematic preference heterogeneity		HMNL model + systematic preference heterogeneity + random opt−out utility	
Utility Function	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value
Alternative-Specific Constant (ASC)								
no vaccination	1.270	< 0.01 ***	0.674	< 0.01 ***	1.070	< 0.01 ***	2.060	0.03 **
Attributes (main effects)								
effectiveness (per 10%)	0.229	< 0.01 ***	0.112	< 0.01 ***	0.001	0.99	0.118	0.01 ***
serious side effects								
1/1.000.000	0.374		0.151		0.060		0.509	
10/1.000.000	0.243	< 0.01 ***	0.116	< 0.01 ***	0.086	0.37	0.037	0.70
100/1.000.000	−0.617	< 0.01 ***	−0.267	< 0.01 ***	−0.146	0.20	−0.546	< 0.01 ***
mild side effects								
1/10	0.026		−0.001		−0.013		0.064	
3/10	0.087	0.22	0.053	0.11	0.112	0.16	0.061	0.42
5/10	−0.113	0.12	−0.053	0.13	−0.099	0.21	−0.125	0.10 *
protection duration								
3 mo	−0.369		−0.147		−0.088		−0.262	
6 mo	0.264	< 0.01 ***	0.112	< 0.01 ***	0.130	0.17	0.081	0.41
12 mo	0.105	0.16	0.035	0.32	−0.042	0.76	0.181	0.06 *
waiting time								
2 wks	−0.045		−0.029		−0.008		0.038	
4 wks	0.045	0.38	0.029	0.21	0.008	0.88	−0.038	0.45
Two-way interactions								
eff x serious10	−0.216	0.08 *	−0.107	0.07 *	−0.102	0.42	−0.028	0.82
eff x serious100	−0.021	0.88	0.005	0.94	−0.263	0.08 *	−0.265	0.06 *
eff x mild3	−0.480	0.71	−0.505	0.38	−1.190	0.37	−0.719	0.58
eff x mild5	−0.296	0.82	0.016	0.98	0.056	0.97	−0.158	0.90
eff x dur6	−0.345	0.01 ***	−0.148	0.02 **	−0.332	0.01 ***	−0.236	0.08 *
eff x dur12	0.065	0.63	0.027	0.66	0.260	0.07 *	0.241	0.08 *
eff x wait4	0.021	0.81	−0.009	0.82	−0.039	0.67	−0.027	0.75
Scale heterogeneity								
deliberative DM style	–		0.338	< 0.01 ***	−0.102	0.16	0.231	0.06 *
high education	–		0.474	< 0.01 ***	0.053	0.38	−0.135	0.12
impact family	–		0.350	0.01 ***	−0.387	< 0.01 ***	−0.451	0.06 *
good health	–		0.433	< 0.01 ***	−0.024	0.76	0.022	0.80
no disease	–		0.230	0.02 **	0.246	< 0.01 ***	0.431	< 0.01 ***
Systematic preference heterogeneity								
age > 64 yr x constant 'no vacc'	–		–		−0.392	0.02 **	−0.003	0.99
age > 64 yr x eff	–		–		−0.052	0.05 **	−0.068	0.02 **
attitude for x constant 'no vacc'	–		–		−1.400	< 0.01 ***	−5.220	< 0.01 ***
attitude for x dur6	–		–		0.057	0.36	0.061	0.31
attitude for x dur12	–		–		0.179	0.02 **	0.017	0.82
attitude for x eff	–		–		0.091	< 0.01 ***	0.005	0.84
attitude for x serious10	–		–		−0.032	0.59	−0.082	0.15
attitude for x serious100	–		–		−0.204	0.01 ***	0.064	0.43
no disease x constant 'no vacc'	–		–		0.563	< 0.01 ***	2.550	< 0.01 ***
deliberative DM style x constant 'no vacc'	–		–		0.922	< 0.01 ***	1.300	< 0.01 ***
deliberative DM style x eff	–		–		0.183	< 0.01 ***	0.120	< 0.01 ***
deliberative DM style x serious10	–		–		0.067	0.38	0.070	0.32
deliberative DM style x serious100	–		–		−0.343	< 0.01 ***	−0.181	0.19
high education x constant 'no vacc'	–		–		1.580	< 0.01 ***	1.030	0.03 **
high education x eff	–		–		0.202	< 0.01 ***	0.284	< 0.01 ***
impact family x constant 'no vacc'	–		–		1.540	0.01 ***	2.140	0.03 **
impact family x eff	–		–		0.153	0.04 **	0.163	0.14
impact family x serious10	–		–		0.179	0.28	0.224	0.21
impact family x serious100	–		–		−0.471	0.04 **	−0.396	0.27
flu symptoms last year x constant 'no vacc'	–		–		−0.454	< 0.01 ***	0.360	0.27
good health x ascn	–		–		0.720	< 0.01 ***	−1.080	0.03 **
good health x eff	–		–		0.146	< 0.01 ***	0.138	< 0.01 ***
last month hosp yes x seriousl0	–		–		0.042	0.51	0.036	0.56
last month hosp yes x seriousl00	–		–		−0.181	0.03 **	−0.177	0.02 **
household alone x eff	–		–		−0.071	< 0.01 ***	−0.102	< 0.01 ***
good health literacy x constant 'no vacc'	–		–		0.514	< 0.01 ***	2.150	< 0.01 ***
good health literacy x dur6	–		–		−0.129	0.05 **	−0.142	0.04 **
good health literacy x dur12	–		–		−0.014	0.86	−0.019	0.81
good health literacy x serious10	–		–		−0.074	0.23	−0.098	0.11
good health literacy x serious100	–		–		0.199	0.01 ***	0.189	0.02 **
good nummeracy x constant 'no vacc'	–		–		−0.266	< 0.01 ***	−0.421	0.22
good numeracy x dur6	–		–		0.106	0.07 *	0.125	0.04 **
good nummeracy x dur12	–		–		−0.039	0.57	−0.060	0.38

(*continued on next page*)

Table 4 (*continued*)

	Model A	Model B	Model C		Model D	
	MNL model	HMNL model	HMNL model + systematic preference heterogeneity		HMNL model + systematic preference heterogeneity + random opt−out utility	
good nummeracy x serious10	–	–	0.071	0.21	0.089	0.10 *
good nummeracy x serious100	–	–	−0.222	< 0.01 ***	−0.256	< 0.01 ***
male x constant 'no vacc'	–	–	0.342	< 0.01 ***	1.350	< 0.01 ***
male x mild3	–	–	−0.007	0.90	−0.027	0.63
male x mild5	–	–	−0.133	0.03 **	−0.085	0.15
no side effects x constant 'no vacc'	–	–	−0.742	< 0.01 ***	1.190	0.01 ***
no side effects x serious10	–	–	0.093	0.13	0.129	0.03 **
no side effects x serious 100	–	–	−0.149	0.05 **	−0.128	0.01 ***
vacc last year x constant 'no vacc'	–	–	−0.759	< 0.01 ***	−4.800	< 0.01 ***
vacc last year x dur6	–	–	0.102	0.11	0.085	0.18
vacc last year x dur12	–	–	0.215	< 0.01 ***	0.171	0.03 **
vacc last year x eff	–	–	0.186	< 0.01 ***	0.148	< 0.01 ***
vacc last year x serious10	–	–	0.040	0.52	0.039	0.52
vacc last year x serious100	–	–	−0.278	< 0.01 ***	−0.084	0.32
flu although being vacc x eff	–	–	−0.063	< 0.01 ***	−0.026	0.33
flu although being vacc x wait4	–	–	0.127	0.04 **	0.104	0.10 *
Random opt-out utility (s.d. of ASC)	–	–	–		5.020	< 0.01 ***
Goodness-of-fit						
LL	−6182	−6140	−4456		−3232	
Number Free Param.	16	21	70		71	
AIC	2.055	2.043	1.501		1.095	
BIC	2.065	2.056	1.546		1.141	
respondents	377	377	377		377	

*p < 0.10; **p < 0.05; ***p < 0.01; MNL = Multinomial Logit; HMNL = Heteroscedastic Multinomial Logit; coeff = coefficient; mo = months; wks = weeks; eff = effectiveness; dur = duration; DM = decision-making; vacc = vaccination; ASC = alternative specific constant; ascn = alternative specific constant opt-out alternative; s.d. = standard deviation; LL = log-likelihood, AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

vaccination, although they did not opt for influenza vaccination in real-life, were quite worried about influenza vaccination, were willing to accept the disadvantages of not being vaccinated, were not supported by the GP to opt for influenza vaccination, lost their invitation for influenza vaccination, and did not have the intention to be vaccinated in the future. Most subjects who said that they would not opt for influenza vaccination, although they opted for influenza vaccination in real-life showed exactly the opposite: they did not worry about influenza vaccination, had a rather positive attitude regarding influenza vaccination, were supported by the GP to opt for influenza vaccination, made an appointment immediately after they received the influenza vaccination invitation, and wanted to be vaccinated in the coming years.

4. Discussion

Our study findings give a high degree of confidence in the results and the external validity of DCE. When choice models took into account both scale and preference heterogeneity, at an aggregate level, the proportion of subjects who opt-in for influenza vaccination was correctly predicted by the DCE. To be more precise, at an aggregate level, the SP choice to opt for influenza vaccination was correctly predicted by a RUM model that took scale and observed preference heterogeneity into account or using an RRM model that took scale, preference heterogeneity and a random opt-out regret into account. At the individual level, both RUM and RRM models with scale, preference heterogeneity and a random opt-out utility/regret had the best prediction ability: up to 91% of individual choices were correctly predicted with a PPV up to 0.96 and NPV up to 0.86. There was 13% discordance between stated choices and actual healthcare utilization. In-depth interviews showed that several dimensions played a role in clarifying this discordance: attitude, social support, action of planning, barriers, and intention.

In a previous DCE study that investigated vaccination uptake (Lambooij et al., 2015), an accuracy of 80% was found between DCE predictions and actual healthcare utilization at an individual level. This is lower than our study finding. According to Lambooij et al. (2015) the PPV and NPV were 0.85 and 0.26, respectively. Our PPV of 0.96 and NPV of 0.86 were substantially higher and more promising. Except for a different target group (hepatitis B vaccination vs. influenza vaccination), a part of the explanation for this phenomenon might be found in the DCE modeling. In contrast to Lambooij et al., our study took scale heterogeneity, observed preference heterogeneity and two-way interactions between the attributes into account, all of which played a significant role in improving the PPV and NPV.

Our in-depth interviews with the discordant-subjects showed a gap between preferences and actions. Although a DCE measures an important component by focusing on relevant attributes that help people decide to participate in influenza vaccination, several other factors played a role in clarifying the discordance between preferences and actions. For respondents that showed discordance, we detected a relative high (conversely, low) amount of presence (absence) of 'worries about influenza vaccination', 'positive attitude regarding influenza vaccination', 'sensitivity for GP's opinion', 'barriers for making an influenza vaccination appointment', and 'how certain the respondent is in his/her decision for influenza vaccination in the (near) future'. Although further research is needed to take these dimensions into account to empirically test how much the inclusion of these omitted variables to our model will improve the external validity, we recommend that DCE practice should change by doing upfront work that emphasizes relevant constraints and inhibitors of the focal behavior, not just what promotes the behavior. For example, presently practice on selecting attributes for DCEs is mostly about what helps people decide to participate. Inhibitors may be among the attributes (e.g., injectable vs oral medication), but it is more likely that inhibitors have to do with exogenous factors like goals, religion, phobias, and social norms. Hence, a DCE measures an important part of the health policy model, but not all.

Our study has several limitations. First, there is the issue of the generalisability of our results. Although the study had a well-considered

Table 5
DCE RRM-results influenza vaccination survey.

	Model E		Model F		Model G		Model H	
	MNL model		HMNL model		HMNL model + systematic preference heterogeneity		HMNL model + systematic preference heterogeneity + random opt-out utility	
Utility Function	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value
Alternative-Specific Constant								
no vaccination	−0.552	< 0.01 ***	−0.138	0.16	−0.823	< 0.01 ***	−1.420	< 0.01 ***
Attributes (main effects)								
effectiveness	−0.015	< 0.01 ***	−0.008	< 0.01 ***	−0.005	0.01 ***	−0.003	0.16
serious side effects								
1/1.000.000	−0.310		−0.108		−0.246		−0.351	
10/1.000.000	−0.186	< 0.01 ***	−0.087	< 0.01 ***	−0.132	0.06 *	−0.032	0.62 ***
100/1.000.000	0.496	< 0.01 ***	0.195	< 0.01 ***	0.378	< 0.01 ***	0.382	< 0.01 ***
mild side effects								
1/10	−0.068		−0.012		−0.023		−0.106	
3/10	−0.053	0.30	−0.039	0.09 *	−0.120	0.05 **	−0.070	0.28
5/10	0.121	0.03 **	0.051	0.05 **	0.143	0.11	0.176	0.01 ***
protection duration								
3 mo	0.275		0.080		0.299		0.313	
6 mo	−0.171	< 0.01 ***	−0.064	0.01 ***	−0.159	0.03 **	−0.104	0.12 **
12 mo	−0.104	0.07 *	−0.015	0.55	−0.140	0.06 *	−0.209	< 0.01 ***
waiting time								
2 wks	−0.011		0.021		−0.068		−0.137	
4wks	0.011	0.80	−0.021	0.30	0.068	0.16	0.137	< 0.01 ***
Two-way interactions								
eff x serious10	0.069	0.57	−0.012	0.82	−0.030	0.79	−0.015	0.89
serious10 x eff	0.069	0.57	0.107	0.05 **	0.088	0.37	−0.074	0.51
eff x serious100	0.058	0.73	0.243	< 0.01 ***	0.223	0.16	0.214	0.17
serious100 x eff	−0.141	0.44	−0.328	< 0.01 ***	−0.046	0.65	0.209	0.03 **
eff x mild3	0.089	0.54	0.085	0.15	0.103	0.38	0.022	0.89
mild3 x eff	−0.070	0.59	−0.054	0.31	−0.021	0.81	0.001	0.96
eff x mild5	−0.195	0.28	−0.071	0.28	−0.231	0.15	−0.252	0.18
mild5 x eff	0.175	0.29	0.074	0.17	0.127	0.18	0.154	0.32
eff x dur6	0.154	0.31	0.027	0.65	0.153	0.23	0.128	0.33
dur6 x eff	0.060	0.64	0.065	0.19	0.108	0.28	0.024	0.83
eff x dur12	−0.083	0.61	−0.106	0.11	0.018	0.90	0.028	0.84
dur12 x eff	0.064	0.65	0.066	0.23	−0.063	0.58	−0.077	0.55
eff x wait4	−0.018	0.08 *	−0.012	0.79	−0.294	0.01 ***	−0.398	< 0.01 ***
wait4 x eff	0.136	0.07 *	0.055	0.09 *	0.210	< 0.01 ***	0.237	< 0.01 ***
Regret Measure								
ξ	2.820	0.07 *	1.430	0.01 ***	100		10	
Scale heterogeneity								
deliberative decision style	–		0.378	< 0.01 ***	−0.249	< 0.01 ***	0.027	0.82
no disease	–		0.224	0.02 **	0.177	< 0.01 ***	0.399	< 0.01 ***
higher education	–		0.490	< 0.01 ***	−0.086	0.10 *	−0.169	0.05 **
impact family	–		0.335	0.01 ***	−0.319	< 0.01 ***	−0.220	0.09 *
good health	–		0.490	< 0.01 ***	−0.184	< 0.01 ***	−0.030	0.73
Systematic preference heterogeneity								
attitude for x constant 'no vacc'	–		–		1.590	< 0.01 ***	3.200	< 0.01 ***
no disease x constant 'no vacc'	–		–		−0.381	< 0.01 ***	−1.710	< 0.01 ***
deliberative DM style x constant 'no vacc'	–		–		−0.580	< 0.01 ***	−0.901	< 0.01 ***
deliberative DM style x eff	–		–		−0.014	< 0.01 ***	−0.010	< 0.01 ***
deliberative DM style x serious10	–		–		−0.153	0.02 **	−0.093	0.10 *
deliberative DM style x serious100	–		–		0.467	< 0.01 ***	0.262	0.03 **
high education x constant 'no vacc'	–		–		−1.050	< 0.01 ***	−0.070	0.80
high education x eff	–		–		−0.015	< 0.01 ***	−0.017	< 0.01 ***
impact family x constant 'no vacc'	–		–		−0.400	< 0.01 ***	−0.373	0.47
flu symptoms last year x constant 'no vacc'	–		–		0.333	< 0.01 ***	−0.001	0.99
last month GP visit x constant 'no vacc'	–		–		−0.103	0.07 *	0.427	0.08 *
good health x constant 'no vacc'	–		–		−0.310	< 0.01 ***	0.791	0.01 ***
good health x eff	–		–		−0.009	< 0.01 ***	−0.008	< 0.01 ***
household alone x eff	–		–		0.004	< 0.01 ***	0.004	< 0.01 ***
good health literacy x constant 'no vacc'	–		–		−0.442	< 0.01 ***	−0.981	< 0.01 ***
good health literacy x dur6	–		–		0.099	0.06 **	0.117	0.01 ***
good health literacy x dur12	–		–		0.004	0.94	−0.012	0.82
good health literacy x mild3	–		–		0.124	0.03 **	0.097	0.07 *
good health literacy x mild5	–		–		−0.068	0.23	−0.038	0.47
good health literacy x seriousl0	–		–		0.058	0.23	0.076	0.09 *
good health literacy x serious100	–		–		−0.152	0.03 **	−0.138	0.03 **
good numeracy x constant 'no vacc'	–		–		0.279	< 0.01 ***	0.428	0.06 *
good numeracy x dur6	–		–		−0.084	0.06 *	−0.089	0.03 **
good numeracy x dur12	–		–		0.038	0.45	0.055	0.23

(*continued on next page*)

Table 5 (*continued*)

	Model E	Model F	Model G		Model H	
	MNL model	HMNL model	HMNL model + systematic preference heterogeneity		HMNL model + systematic preference heterogeneity + random opt-out utility	
good numeracy x serious10	–	–	−0.051	0.22	−0.063	0.10 *
good numeracy x serious100	–	–	0.154	0.01 ***	0.152	0.01 ***
male x constant 'no vacc'	–	–	−0.239	< 0.01 ***	−0.813	< 0.01 ***
male x mild3	–	–	−0.007	0.88	0.023	0.58
male x mild5	–	–	0.153	0.01 ***	0.080	0.08 *
vacc last year x constant 'no vacc'	–	–	1.140	< 0.01 ***	3.350	< 0.01 ***
vacc last yearx dur6	–	–	−0.085	0.07 *	−0.074	0.09 *
vacc last year x dur12	–	–	−0.160	< 0.01 ***	−0.108	0.04 **
vacc last year x eff	–	–	−0.006	< 0.01 ***	−0.007	< 0.01 ***
vacc last year x serious10	–	–	−0.020	0.65	−0.022	0.60
vacc last year x serious100	–	–	0.145	0.04 **	0.034	0.62
flu although being vaccinated x eff	–	–	0.005	< 0.01 ***	0.002	0.30
flu although being vaccinated x wait4	–	–	−0.107	0.02 **	−0.058	0.19
no side effects x constant 'no vacc'	–	–	0.565	< 0.01 ***	−0.525	0.05 **
no side effects x eff	–	–	0.014	< 0.01 ***	−0.044	0.99
no side effects x serious10	–	–	−0.075	0.09 *	−0.082	0.05 **
no side effects x serious100	–	–	0.135	0.03 **	0.092	0.12
Random opt-out regret (s.d. of ASC)	–	–	–		3.240	< 0.01 ***
Goodness-of-fit						
LL	−6144	−6094	−4408		−3267	
Number Free Param.	24	29	69		70	
AIC	2.045	2.030	1.484		1.106	
BIC	2.061	2.049	1.529		1.152	
respondents	377	377	377		377	

*p < 0.10; **p < 0.05; ***p < 0.01; MNL = Multinomial Logit; HMNL = Heteroscedastic Multinomial Logit; coeff = coefficient; mo = months; wks = weeks; eff = effectiveness; dur = duration; DM = decision-making; vacc = vaccination; ASC = alternative specific constant; ascn = alternative specific constant opt-out alternative; s.d. = standard deviation; LL = log-likelihood, AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Table 6
DCE model fit and prediction results on an aggregate and individual level.

		Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H
		Random Utility Maximization				Random Regret Minimization			
		MNL	HMNL	HMNL+	HMNL++	MNL	HMNL	HMNL+	HMNL++
Aggregate level									
Observed vaccination uptake based on revealed **(RP)** data		**65.0%**							
Observed vaccination uptake based on choice task **(SP)** mimicking the real-world choice of influenza vaccination		**68.4%**							
Predicted vaccination uptake based on DCE:	mean	**66.4%**	**62.1%**	**66.9%**	**63.8%**	**74.4%**	**63.9%**	**77.9%**	**68.5%**
	lower bound CI	66.4%	61.6%	63.4%	60.9%	74.4%	63.4%	75.1%	65.8%
	upper bound CI	66.4%	62.5%	70.4%	66.6%	74.4%	64.5%	80.7%	71.2%
Individual level									
Proportion of individuals choices that were **correctly predicted** at an **individual level**		56.1%	53.9%	80.0%	90.8%	59.0%	54.5%	79.1%	91.1%
Positive Predictive Value (PPV)		0.66	0.62	0.84	0.93	0.63	0.58	0.81	0.96
Negative Predictive Value (NPV)		0.34	0.37	0.71	0.86	0.38	0.42	0.77	0.80
LogLikelihood		−6182	−6140	−4456	−3232	−6144	−6094	−4408	−3267
Degrees of freedom		16	21	70	71	24	29	69	70
AIC		2.055	2.043	1.501	1.095	2.045	2.030	1.484	1.106
BIC		2.065	2.056	1.546	1.141	2.061	2.049	1.529	1.152
Respondents (n)		377							

Note: MNL = multinomial model; HMNL = heteroscedastic model; HMNL+ = heteroscedastic model plus systematic preference heterogeneity; HMNL++ = Heteroscedastic model plus systematic preference heterogeneity plus random intercept; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

and systematic study approach, the results were based on one medical context only due to practical constraints. Further research is therefore needed to determine whether our results hold in other medical contexts as well. Our study design can be used as a guide for future studies. Second, due to the systematic approach we used starting from common practice in health economics, other choice models (e.g., latent class model, mixed logit model) and ensemble analyses were not investigated. Further research is therefore recommended, although DCE researchers should be cautious about over-fitting their data (as showed by our study: a better model fit will not automatically lead to a better prediction). Third, the respondents' vaccination behaviour might have impacted their responses during the interviews. That is, these respondents may have 'defended' their real behaviour choice (i.e., cognitive discordance). This might have biased the interview results. It would be helpful in future studies to collect information on the theoretical factors of the I-Change model prior to the health behaviour as

Table 7
Interview results of discordance between SP and actual healthcare utilization.

	Said yes, did no (N = 8)	Said yes, did yes (N = 4)	Said no, did yes (N = 11)	Said no, did no (N = 1)
Cognizanze - Awareness				
1 Aware of flu vaccination (yes)	8	4	11	1
2 Aware s/he belongs to the target group (yes)	8	4	8	1
3 Feels s/he belongs to the target group (yes)	7	4	8	1
4 Knows no out-of-pocket money needed (yes)	6	4	7	0
Knowledge				
5 About influenza incidence (yes)	7	2	10	1
6 About different types of influenza (yes)	8	4	11	1
7 About possibility getting flu although vaccinated (yes)	8	4	11	1
Attitude				
8 Positive attitude re flu vaccination (yes)	**1**	**4**	**9**	1
9 Worries about flu vaccination (yes)	**6**	**0**	**2**	0
10 Prepared to take the risk of not being vaccinated (yes)	**7**	**0**	7	**0**
Self-efficacy				
11 Situations where flu vaccination is difficult (no)	7	4	11	1
Social support				
12 Influenced by family/friends (yes)	4	3	3	0
13 Thinks support is important (no)	7	3	11	1
14 Did GP support flu vaccination (yes)	**2**	2	**7**	**1**
15 Is the GP opinion important for him/her (yes)	**7**	1	**9**	**1**
Action - Planning				
16 S/he made an appointment immediately	**0**	**4**	**10**	0
Barriers				
17 Received invitation (yes)	8	4	11	1
18 Lost invitation; s/he did not make an appointment (yes)	**8**	**4**	**7**	0
19 Knows where to get the flu vaccination (yes)	8	4	11	1
Intention				
20 Wants to be flu vaccinated next year (yes or definitely yes)	**0**	**4**	**11**	0
21 Wants to be flu vaccinated every year (yes or definitely yes)	**0**	**4**	**11**	0

well. Fourth, although the response rate was a bit higher than a priori expected, a response rate of 31.2% is still not ideal. As a result there might be a probable lack of representativeness of the individuals who responded and completed the survey. Finally, a mere-measurement effect may exist in our sample as the percentage of respondents (65%) who actually opt for flu vaccination seemed to be higher (although within the 95% con-fidence interval range) than current Dutch practice for the same age group (58%; CI 51%–65%).

In conclusion, our study shows that a DCE does its job of predicting actual healthcare choices of respondents if at least scale and preference heterogeneity are taken into account, but that there is a gap between preferences and actions. DCEs are typically focused on the attributes and the resulting evaluation of a healthcare intervention. In many cases, including vaccination programs, there are factors inhibiting individuals from acting in concordance with their preferences. Follow-up interviews with patients showed that such inhibitors are often not directly related to the intervention, as they related to goals, religion, phobias, and social norms. Doing upfront work that emphasizes constraints/inhibitors of the focal behavior, not just what promotes the behavior, might further improve predictive ability. Further research is recommended to determine whether our conclusion holds for predicting healthcare choices of an external sample.

Ethics approval

Approval for the study was obtained from the Medical Ethics Committee, Erasmus MC (MEC- 2016-095).

Acknowledgement

We thank all respondents for their great contributions, and the GPs Patrick Bindels, Sylvia Buis and Jan Huisman who provided insights and help that greatly assisted the research. We would also like to show our gratitude to Marcel Jonker for his contribution to the DCE design, Hein de Vries for his great help to the interview guide, and Samare Huls and Laura Koelemeijer for conducting the interviews. Grant support was 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, the collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the article for publication.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. J. Environ. Econ. Manag. 26 (3), 271–292. https://doi.org/10.1006/jeem.1994.1017.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. https://doi.org/10.1016/0749-5978(91)90020-T.

Bandura, A., 1986. In: Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall, Inc., Englewood Cliffs, NJ, US. https://doi.org/10.1037/13273-005.

Bhat, C.R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. Transp. Res. Part B Methodol. 29 (6), 471–483. https://doi.org/10.1016/0191-2615(95)00015-6.

Bierlaire, M., 2016. Estimating Choice Models with Latent Variables with PythonBiogeme, Technical Report TRANSP-OR 160628. Transport and Mobility Laboratory. ENAC, EPFL.

Bliemer, M., Rose, J., 2011. Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transp. Res. Part A Policy Pract 45, 63–79.

Bliemer, M., Rose, J., Chorus, C., 2014. Dominancy in Stated Choice Surveys and its Impact on Scale in Discrete Choice Models.

Bliemer, M., Rose, J., Chorus, C., 2017. Detecting dominancy in stated choice data and accounting for dominancy-based scale differences in logit models. Transp. Res. Part B Methodol. 102, 83–104.

Brownstone, D., Train, K., 1998. Forecasting new product penetration with flexible substitution patterns. J. Econom. 89 (1–2), 109–129. https://doi.org/10.1016/S0304-4076(98)00057-8.

Burns, V.E., Ring, C., Carroll, D., 2005. Factors influencing influenza vaccination uptake in an elderly, community-based sample. Vaccine 23 (27), 3604–3608. https://doi.org/10.1016/j.vaccine.2004.12.031.

Cameron, T.A., Poe, G.L., Ethier, R.G., Schulze, W.D., 2002. Alternative non-market value-elicitation methods: are the underlying preferences the same? In: Journal of Environmental Economics and Management, vol. 44. pp. 391–425. https://doi.org/10.1006/jeem.2001.1210.

Carlsson, F., Martinsson, P., 2001. Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. J. Environ. Econ. Manag. 41 (2), 179–192. https://doi.org/10.1006/jeem.2000.1138.

Cheraghi-Sohi, S., Hole, A., Mead, N., et al., 2008. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann. Fam. Med. 6 (2), 107–115.

Clark, M., Determann, D., Petrou, S., Moro, D., de Bekker-Grob, E.W., 2014. Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics 32 (9), 883–902. https://doi.org/10.1007/s40273-014-0170-x.
de Bekker-Grob, E., Essink-Bot, M., Meerding, W., Pols, H., Koes, B., Steyerberg, E., 2008. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos. Int. 19 (7), 1029–1037.
de Bekker-Grob, E., Hol, L., Donkers, B., et al., 2010. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. Value Health 13 (2), 315–323.
de Bekker-Grob, E.W., Donkers, B., Jonker, M.F., Stolk, E.A., 2015. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. Patient 8 (5), 373–384. https://doi.org/10.1007/s40271-015-0118-z.
de Bekker-Grob, E.W., Veldwijk, J., Jonker, M., et al., 2018. The impact of vaccination and patient characteristics on influenza vaccination uptake of elderly people: a discrete choice experiment. Vaccine 36 (11), 1467–1476. https://doi.org/10.1016/j.vaccine.2018.01.054.
de Bekker-Grob, E.W., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. Health Econ. 21 (2), 145–172. https://doi.org/10.1002/hec.1697.
de Bekker-Grob, E., Swait, J., Kassahun, H., et al., 2019. Are healthcare choices predictable? the impact of discrete choice experiment designs and models. Value Heal 22 (9), 1050–1062.
de Vries, H., 2017. An integrated approach for understanding health behavior; the I-change model as an example. Psychol. Behav. Sci. Int. J. 2 (2), 555–585.
de Vries, H., Mesters, I., van de Steeg, H., Honing, C., 2005. The general public's information needs and perceptions regarding hereditary cancer: an application of the Integrated Change Model. Patient Educ. Couns. 56 (2), 154–165.
Delbecq, A.L., Van de Ven, A.H., 1971. A group process model for problem identification and program planning. J. Appl. Behav. Sci. https://doi.org/10.1177/002188637100700404.
Determann, D., Korfage, I.J., Lambooij, M.S., et al., 2014. Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. PLoS One 9 (7). https://doi.org/10.1371/journal.pone.0102505.
Determann, D., Korfage, I., Fagerlin, A., et al., 2016. Public preferences for vaccination programmes during pandemics caused by pathogens transmitted through respiratory droplets – a discrete choice experiments in four European countries. Euro Surveill. 22 (21).
Fifer, S., Rose, J., Greaves, S., 2014. Hypothetical bias in Stated Choice Experiments: is it a problem? And if so, how do we deal with it? Transp. Res. Part A Policy Pract 61, 164–177. https://doi.org/10.1016/j.tra.2013.12.010.
Goodwin, K., Viboud, C., Simonsen, L., 2006. Antibody response to influenza vaccination in the elderly: a quantitative review. Vaccine 24, 1159–1169. https://doi.org/10.1016/j.vaccine.2005.08.105.
Hensher, D., Rose, J., Greene, W., 2015. Applied Choice Analysis, second ed. Cambridge University Press, Cambridge, pp. 502 see.
Hess, S., 2010. Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. J. Choice Model 3 (2), 134–152.
Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. J. Choice Model. https://doi.org/10.1016/j.jocm.2017.03.001.
Janssen, E., Marshall, D., Hauber, A., Bridges, J., 2017. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev. Pharmacoecon. Outcomes Res. 17 (6), 531–542.
Janz, N.K., Becker, M.H., 1984. The health belief model: a decade later. Health Educ. Behav. https://doi.org/10.1177/109019818401100101.
Johnson, R., Orme, B., Johnson, R., Orme, B., 2003. Getting the Most from CBC. Sawtooth Softw Res Pape ser Sawtooth Software, Sequim.
Kesternich, I., Heiss, F., McFadden, D., Winter, J., 2013. Suit the action to the word, the word to the action: hypothetical choices and real decisions in Medicare Part D. J. Health Econ. 32 (6), 1313–1324. https://doi.org/10.1016/j.jhealeco.2012.11.006.
Kohler, R., Lee, C., Gopal, S., Reeve, B., Weiner, B., Wheeler, S., 2015. Developing a discrete choice experiment in Malawi: eliciting preferences for breast cancer early detection services. Patient Prefer. Adherence 9, 1459–1472.
Lambooij, M.S., Harmsen, I.A., Veldwijk, J., et al., 2015. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Med. Res. Methodol. 15 (1). https://doi.org/10.1186/s12874-015-0010-5.
Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. PharmacoEconomics 26 (8), 661–677. https://doi.org/10.2165/00019053-200826080-00004.
Lancsar, E., Swait, J., 2014. Reconceptualising the external validity of discrete choice experiments. PharmacoEconomics 32 (10), 951–965. https://doi.org/10.1007/s40273-014-0181-7.
Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am. Psychol. https://doi.org/10.1037/0003-066X.57.9.705.
Locke, E.A., Latham, G.P., 2006. New directions in goal-setting theory. Curr. Dir. Psychol. Sci. https://doi.org/10.1111/j.1467-8721.2006.00449.x.
Luce, M.F., Payne, J.W., Bettman, J.R., 1999. Emotional trade-off difficulty and choice. J. Mark. Res. 36 (2), 143. https://doi.org/10.2307/3152089.
Mahieu, P., Andersson, H., Beaumais, O., Crastes Sourd, R., Hess, S., Wolff, F., 2017. 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. 98 (3), 201–220.
Mark, T.L., Swait, J., 2004. Using stated preference and revealed preference modeling to evaluate prescribing decisions. Health Econ. 13 (6), 563–573. https://doi.org/10.1002/hec.845.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Front. Econ. 1 (2), 105–142. https://doi.org/10.1108/eb028592.
Mueller, S., Osidacz, P., Francis, I.L., Lockshin, L., 2010. Combining discrete choice and informed sensory testing in a two-stage process: can it predict wine market share? Food Qual. Prefer. 21 (7), 741–754. https://doi.org/10.1016/j.foodqual.2010.06.008.
Natter, M., Feurstein, M., 2002. Real world performance of choice-based conjoint models. Eur. J. Oper. Res. 137 (2), 448–458. https://doi.org/10.1016/S0377-2217(01)00147-3.
Orme, B., 1998. Sample Size Issues for Conjoint Analysis Studies.
Prochaska, J.O., Diclemente, C.C., 1986. Toward a comprehensive model of change. In: Treating Addictive Behaviors, https://doi.org/10.1007/978-1-4613-2191-0_1.
Quaife, M., Terris-Prestholt, F., Di Tanna, G., Vickerman, P., 2018. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur. J. Health Econ.
Revelt, D., Train, K., 2000. Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier. https://cloudfront.escholarship.org/dist/prd/content/qt1900p96t/qt1900p96t.pdf.
Rowen, D., Mulhern, B., Stevens, K., Vermaire, J., 2018. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. Value Health 21 (10), 1234–1242.
Ryan, M., 2004. Discrete choice experiments in health care. BMJ. https://doi.org/10.7326/0003-4819-96-1-125_1.
Ryan, M., Bate, A., Eastmond, C.J., Ludbrook, A., 2001. Use of discrete choice experiments to elicit preferences. Qual. Health Care 10 (Suppl. I), i55–i60. https://doi.org/10.1136/qhc.0100055.
Sadique, M.Z., Devlin, N., Edmunds, W.J., Parkin, D., 2013. The effect of perceived risks on the demand for vaccination: results from a discrete choice experiment. PLoS One 8 (2). https://doi.org/10.1371/journal.pone.0054149.
Salisbury, L.C., Feinberg, F.M., 2010. Alleviating the constant stochastic variance assumption in decision research: theory, measurement, and experimental test. Mark. Sci. 29 (1), 1–17. https://doi.org/10.1287/mksc.1080.0464.
Sándor, Z., Wedel, M., 2005. Heterogeneous conjoint choice designs. J. Mark. Res. 42 (2), 210–218. https://doi.org/10.1509/jmkr.42.2.210.62285.
Scarpa, R., Ferrini, S., Willis, K., 2005. Performance of error component models for status-quo effects in choice experiments. In: Applications of Simulation Methods in Environmental and Resource Economics, pp. 247–273 1571-487X.
Shono, A., Kondo, M., 2014. Parents' preferences for seasonal influenza vaccine for their children in Japan. Vaccine 32 (39), 5071–5076. https://doi.org/10.1016/j.vaccine.2014.07.002.
Soekhai, V., de Bekker-Grob, E., Ellis, A., Vass, C., 2019. Discrete choice experiments in health economics: past, present and future. PharmacoEconomics 37 (2), 201–226.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, MA.
van Cranenburgh, S., Guevara, C., Chorus, C., 2015. New insights on random regret minimization models. Transp. Res. Part A Policy Pract 74, 91–109.
Watson, V., Ryan, M., Watson, E., 2009. Valuing experience factors in the provision of chlamydia screening: an application to women attending the family planning clinic. Value Health 12 (4), 621–623. https://doi.org/10.1111/j.1524-4733.2008.00451.x.
Watson, V., Becker, F., de Bekker-Grob, E., 2017. Discrete choice experiment response rates: a meta-analysis. Health Econ. 26 (6), 810–817.