The Patient - Patient-Centered Outcomes Research (2023) 16:641–653
https://doi.org/10.1007/s40271-023-00643-w

ORIGINAL RESEARCH ARTICLE

Maximum Acceptable Risk Estimation Based on a Discrete Choice Experiment and a Probabilistic Threshold Technique

Jorien Veldwijk[1,2,3] · Rachael Lynn DiSantostefano[4] · Ellen Janssen[4] · Gwenda Simons[5] · Matthias Englbrecht[6,7] · Karin Schölin Bywall[8] · Christine Radawski[9] · Karim Raza[5,10,11] · Brett Hauber[12,13] · Marie Falahee[5]

Accepted: 19 July 2023 / Published online: 30 August 2023
© The Author(s) 2023

Abstract

Objective We aimed to empirically compare maximum acceptable risk results estimated using both a discrete choice experiment (DCE) and a probabilistic threshold technique (PTT).

Methods Members of the UK general public ($n = 982$) completed an online survey including a DCE and a PTT (in random order) measuring their preferences for preventative treatment for rheumatoid arthritis. For the DCE, a Bayesian D-efficient design consisting of four blocks of 15 choice tasks was constructed including six attributes with varying levels. The PTT used identical risk and benefit attributes. For the DCE, a panel mixed-logit model was conducted, both mean and individual estimates were used to calculate maximum acceptable risk. For the PTT, interval regression was used to calculate maximum acceptable risk. Perceived complexity of the choice tasks and preference heterogeneity were investigated for both methods.

Results Maximum acceptable risk confidence intervals of both methods overlapped for serious infection and serious side effects but not for mild side effects (maximum acceptable risk was 32.7 percent-points lower in the PTT). Although, both DCE and PTT tasks overall were considered easy or very easy to understand and answer, significantly more respondents rated the DCE choice tasks as easier to understand compared with those who rated the PTT as easier (7-percentage point difference; $p < 0.05$).

Conclusions Maximum acceptable risk estimate confidence intervals based on a DCE and a PTT overlapped for two out of the three included risk attributes. More respondents rated the DCE as easier to understand. This may suggest that the DCE is better suited in studies estimating maximum acceptable risk for multiple risk attributes of differing severity, while the PTT may be better suited when measuring heterogeneity in maximum acceptable risk estimates or when investigating one or more serious adverse events.

1 Introduction

There is increasing interest in the use of information about the benefit-risk trade-offs that people or patients make when considering treatment among decision makers at all stages of medicine development, regulatory approval, and reimbursement [1–3]. In patient-centered outcomes research, risk is the chance of a negative outcome, such as developing rheumatoid arthritis (RA). Treatments may reduce some risks (i.e., benefits) and induce other risks, such as side effects. The maximum level of risk that people are willing to accept in exchange for a given increase in benefit (also known as the maximum acceptable risk [MAR]) is one of the measures that can be used by pharmaceutical companies and regulatory agencies to support, for instance, target product profile development, endpoint selection, benefit-risk assessment, and regulatory approval.

Different methods can be used to estimate MAR for a given medicinal product, with discrete choice experiments (DCEs) being the most widely applied method [4]. Discrete choice experiments ask respondents to evaluate full treatment profiles against each other. It is assumed that respondents consider all treatment attributes included in these profiles and select the profile that provides them with the highest utility. Discrete choice experiments have been applied broadly throughout healthcare settings [5] and the first insights into external validity in healthcare have been published [6–10]. However, DCEs are relatively resource intense to administer and the cognitive burden on respondents is considered by some to be relatively high [11, 12]. This method further requires complex experimental designs [13], statistical modeling techniques [14], and relatively large sample sizes [15]. Finally, MAR is a secondary outcome measure in DCE studies, which is indirectly calculated

Brett Hauber and Marie Falahee shared last authorship.

Extended author information available on the last page of the article

Key Points

Although they elicit preferences and risk tolerance differently, discrete choice experiments and the probabilistic threshold technique lead to similar conclusions with regard to the general magnitude of maximum acceptable risk for serious adverse events.

Discrete choice experiments seem better suited in studies that aim to estimate the maximum acceptable risk for multiple risk attributes of differing severity.

The probabilistic threshold technique is likely better suited to measuring heterogeneity in the maximum acceptable risk estimates across a relatively small population, particularly for serious adverse events.

More respondents perceived the choice tasks of a discrete choice experiment to be easy or very easy to understand compared with those who rated the choice tasks of a probabilistic threshold technique the same way.

from estimated model output; as a result, the ability to identify covariates that are associated with heterogeneity in MAR is expected to be limited [16].

Other preference methodologies can be used to estimate MAR, including the probabilistic threshold technique (PTT) [16]. A recent review identified over 43 applications of this method across the healthcare sector [16] and additional applications of the PTT have been used to support regulatory decisions [17, 18]. The PTT directly measures MAR by asking respondents to choose between two treatment profiles, of which one represents an improved benefit over the other, but also imposes additional risk. Subsequently, the level of risk in the profile providing the beneficial treatment is varied until the respondent switches their choice to the alternative profile. In contrast to the DCE, this method estimates the threshold value directly and at an individual level, does not require a complex experimental design, and is typically applied in smaller samples [16].

Although both methods are able to produce MAR estimates, research comparing the extent to which results from a DCE and a PTT differ when applied to the same research question is lacking. Because both methods can be used to determine MAR, applying them to a similar research question should ideally result in MAR estimates with overlapping confidence intervals. To date, there is no evidence-based guidance on what points to consider when selecting one of these methods over the other for the purpose of eliciting preferences and estimating MAR. Therefore, the current study included DCE and PTT exercises in ways these methods have been applied previously to answer the same clinical research question. The study aimed to empirically compare the results of the DCE and the PTT for estimating MAR.

2 Methods

2.1 Case Study and Ethical Approval

Part of a study assessing preferences of the general public and first-degree relatives of patients with rheumatoid arthritis (RA) for the preventative treatment of RA was used as a case study for this research. As stated before, the DCE and PTT choice tasks in this study were designed, conducted, and analyzed individually, with the DCE following good research practices [14, 19–22] and the PPT following the approach used in a recent empirical study [23]. That is, the DCE was used to estimate an indirect utility function and then calculate MAR using the resulting estimates while the PTT was used to estimate MAR directly. Even though one method is applied in utility space while the other is applied in MAR space, they are both ultimately used to estimate the same measure of risk tolerance when applied to the same research question and both have been used to inform previous regulatory decisions. The study is described in detail elsewhere [24–26].

2.2 Attribute and Level Selection

A literature review [27] followed by a qualitative study using focus groups and individual interviews including a ranking exercise were conducted to inform attribute selection [24–26]. Selection of final attributes was agreed by an international multidisciplinary team of clinical researchers, patient preference experts, and patient research partners. Attribute levels were selected based on current clinical evidence and expert consultation. Table 1 includes a full overview of all attributes and levels included in the study.

2.3 Experimental Design DCE

Ngene software (ChoiceMetrics. Sydney, NSW, Australia) was used to construct a Bayesian D-efficient design [28]. Prior information on the attribute-level parameters was based on the previous literature [27] and best guesses for a pilot study and outcomes of an initial analysis (conditional logit) of pilot data ($n = 100$) for the main survey. For both the pilot and the final design, 60 unique choice tasks were generated, which were divided into four blocks. In the DCE, participants were randomized to one of the blocks, so each

participant answered only 15 choice tasks. The choice tasks each consisted of two hypothetical treatment scenarios as generated in the experimental design and a no-treatment option. Within each block, choice tasks were shown in a fixed order to respondents, but the order of the treatment alternatives (i.e., being the left or right alternative) was randomized. The no-treatment options were always presented on the far-right side of the choice task. The design was restricted to exclude the following combinations of attribute levels in treatment profiles: taking a pill every month or every 6 months, having an injection daily, or having a drip daily or weekly. Additionally, interactions between effectiveness and the chance of a serious infection and between effectiveness and the chance of serious side effects were included in the design. The full experimental DCE design can be found in the Electronic Supplementary Material (ESM).

2.4 Design of the PTT

The PTT has a similar choice task format as the DCE and was developed based on the final DCE design [26]. In the first question, referred to as the baseline question, participants were asked to choose between no treatment (a 60% chance of developing RA and no increased safety risks) or treatment (fixed benefit and varying levels of increased risks). The fixed benefit in the PTT was a 40-percentage point reduction in the chance of developing RA, which corresponded to the difference in the highest and lowest levels of this attribute in the DCE. Table 1 shows the starting values of each of the risk attributes for the PTT. After answering the baseline question, participants were asked a series of follow-up questions. In each follow-up question, the 'no treatment' alternative remained fixed. In the treatment alternative, the level of benefit remained fixed, and the level of one of the risk attributes (i.e., mild side effects, serious infection, or serious side effects) was varied systematically while the other two risks were fixed at the baseline level. For the risk that varied, the subsequent levels were either lower or higher than the baseline risk in the treatment alternative depending on the baseline response. A maximum of two follow-up questions were asked after the baseline for each included risk attribute. If respondents indicated they would accept the highest level of risk included in the PTT exercise, an open-ended question was included to ask respondents to indicate the highest risk they were willing to accept. Whenever respondents did not accept the lowest risk included in the PTT exercise, a follow-up question asked if they were (yes or no) willing to accept any risk at all. This procedure

Table 1 Attributes and levels selected for the DCE and the PTT[a]

Attributes	Levels
Chance of developing RA reduced from 60% to	10% or 10 in 100 20% or 20 in 100 30% or 30 in 100 40% or 40 in 100 60% or 60 in 100 (included in opt-out only)
How the treatment is taken	A shallow injection under the skin A drip into the vein One or two tablets Not applicable (included in opt-out only)
How often the medication has to be taken	Daily Weekly Monthly Every 6 months Not applicable (included in opt-out only)
Chance of mild side effects	2% or 2 in 100 people 5% or 5 in 100 people 10% or 10 in 100 people None or 0 in 100 people (included in opt-out only)
Chance of a serious infection due to treatment	0% or 0 in 100 people 1% or 1 in 100 people 5% or 5 in 100 people None or 0 in 100 people (included in opt-out only)
Chance of a serious side effect	0.001% or 1 in 100,000 people 0.02% or 20 in 100,000 people 0.1% or 100 in 100,000 people None or 0 in 100,000 people (included in opt-out only)

DCE discrete choice experiment, *PTT* probabilistic threshold technique, *RA* rheumatoid arthritis

[a]Starting values for the risks in the PTT were 5%, 2%, and 0.02% for mild side effects, serious infection, and serious side effects, respectively

was repeated for each risk attribute to define the risk intervals for each participant for each risk presented in the survey. The order in which respondents completed the PTT questions for the three risk attributes was randomized. See the ESM for an example of the design of the PTT.

2.5 Survey

The online survey, programmed in SurveyEngine GmbH software, consisted of six sections and was developed in close collaboration with patient research partners to enhance content and accessibility. First, respondents completed informed consent, and a demographic questionnaire. Second, they were asked to read a description of RA and risk factors for RA. This information was followed by comprehension questions to test the participant's understanding of the information presented. Participants were then asked to imagine they had started to develop joint pain and had received test results that indicated they had a 60% chance of developing RA in the following 2 years. Subsequently, attributes and levels were explained in detail, and examples of choice tasks were given, including a walk-through example. In the third part, either the DCE or the PPT choice task questions (participants were randomly assigned to either receive the DCE or the PTT first) were presented. The effectiveness attribute and all risk attributes were displayed using percentages to increase comparability. In addition, frequencies were added to these choice tasks to enhance interpretation and participants could view the explanation of the attribute and the levels (including icon arrays for the risk-related levels) using pop-up windows during the completion of the choice task. Fourth, participants were asked to complete the Single Item Literacy Screener (SILS [29]) and the 3-Item Version of the Subjective Numeracy Scale (SNS-3 [30]). Fifth, participants were shown a walk-through for the choice tasks of the second method, followed by the actual choice tasks. Sixth, participants were asked to complete further measures of psychological constructs, including their family history of RA.

After completing each preference method, participants were asked to complete two feedback questions related to how easy it was to understand and answer the choice questions. A copy of the full survey is included in the ESM.

2.6 Pre-Pilot and Pilot Testing

The survey was pre-tested in a convenience sample ($N = 15$) of members from the general public, first-degree relatives of patients with RA, and patient research partners in the UK using qualitative think-aloud interviews and written feedback forms, to test the online survey system, remove software bugs, change the wording if necessary, and ensure adequate interpretation of risk information presented in the choice tasks. These participants were paid £20 (i.e., approximately $US27) in shopping vouchers. To optimize the statistical efficiency of the final DCE design, a survey pilot was conducted with 100 members of the general public in the UK.

2.7 Participant Recruitment

Participants were recruited via an online survey panel of members of the general population in UK provided by Survey Engine GmbH. Eligible participants were 18 years of age or older, without a diagnosis of RA [24]. Potential participants received an e-mail invitation to take part in the survey with a unique password-protected link to the online survey itself. After completing the survey, panel members were credited with panel points (equivalent to approximately £2 [i.e., approximately US$2.75] for a 30-minute online survey) and additional sources of information and support about RA and risk factors for RA. Recruitment was continued until 1000 completed surveys were returned [24].

2.8 Data Analysis

Only completed surveys were included in the analysis. Discrete choice experiment analyses were conducted in Nlogit 6 (Econometric Software Inc., Plainview, New York, USA) and the PTT analyses were conducted in R Statistical Software 3.6.1 (R Core Team, Vienna, Austria). All surveys that were completed in less than 5 minutes were excluded. The data from all participants who took between 5 and 9 minutes to answer the survey AND answered two out of the three comprehension questions incorrectly AND showed flatlining behavior in the DCE tasks were excluded because of suspicion of low-quality data. Results were considered statistically significant if $p < 0.05$.

2.8.1 Respondent Feedback Questions

Frequencies were calculated and chi-square tests were conducted to compare respondents' ratings of the perceived difficulty to understand and answer the DCE and PTT questions.

2.8.2 DCE Analysis

Panel random parameter logit (RPL) models were constructed to adjust for the multi-level structure of the data in order to be able to correct for preference heterogeneity [14]. The linearity of all non-categorical attributes was assessed visually by plotting the coefficients and based on a dummy variable in a spline function. Variables were included as either linear or categorical (using effects coding [31, 32])

in the utility function, except for effectiveness, for which linearity was assumed to facilitate MAR calculations. The contributions to model goodness of fit of interaction terms that were pre-specified in the experimental design were tested (i.e., effectiveness-serious infection and effectiveness-serious side effects). Interaction terms that significantly contributed to the model fit (Log likelihood ratio test) were included. To test whether the order in which participants answered the choice tasks (i.e., DCE first or PTT first) influenced the outcomes, a variable that identified that order was included in the model as an interaction term with the attribute levels.

All attributes were included as random parameters. Based on a model fit test, it was determined what distribution should be assumed for those parameters (normal, triangular, or lognormal). The systematic utility component (V), which describes the measurable utility of a specific treatment based on the attributes included in the DCE, was tested using the equation below. The β_0 represents the alternative specific constant for the opt-out alternative, and other estimates are the attribute-level estimates that indicate the relative importance of each attribute level.

$V_{\text{alternative A|B}} = \beta 1_i$ * chance of developing RA + $\beta 2_i$ * mode of administration $_{\text{injection}}$ + $\beta 3_i$ * mode of administration $_{\text{drip}}$ + $\beta 4_i$ * frequency of administration + $\beta 5_i$ * chance of mild side effect $_{5\%}$ + $\beta 6_i$ chance of mild side effect $_{10\%}$ + $\beta 7_i$ * chance of serious infection + $\beta 8_i$ * chance of serious side effects $_{20 \text{ in } 100.000}$ + $\beta 9_i$ * chance of serious side effects $_{1 \text{ in } 100.000}$ + $\beta 10_i$ * effectiveness * chance of serious infection

$V_{\text{opt-out}} = \beta 0_i$

Maximum acceptable risk was calculated in all instances where the risk attribute was linearly coded following the equation below. When a risk attribute was effects coded, linearity was assumed between the two highest risk levels; this was further extrapolated beyond the levels included in the DCE experiment when necessary. Calculations were made both based on aggregate average attribute-level estimates as well as individual attribute-level estimates retrieved from the RPL model.

$$MAR = -\frac{(\beta_{\text{Chance of developing RA}})}{(\beta_{k=risk\ attribute})}$$

2.8.3 PTT Analysis

For the PTT, the series of threshold technique questions resulted in a threshold interval representing the risk level each participant was willing to accept in exchange for the benefit (reduction in the chance of developing RA). These data were explored using descriptive statistics. The proportion of respondents selecting each of the risk levels as the MAR were displayed. Whenever respondents indicated a MAR beyond the pre-set range in the open-ended follow-up question, these risks were grouped into intervals of 10 percentage points (meaning that, e.g., a response of 23% would fall in the 20–30% interval). The formal analyses were conducted using interval regression models, in which the data were interval censored because the threshold falls within an interval with fixed endpoints. An interval regression model was fitted using a Tobit model to account for the fact that the interval has both a fixed upper bound (being the highest acceptable risk as indicated by respondents in the open-ended risk question), resulting in left-censored data, and a fixed lower bound, resulting in right-censored data [33]. For each risk attribute, ($r \in$ [*mild side effects*, *serious infection*, *serious side effects*]), the risk interval threshold, *Threshold*, was regressed on the variable that indicated whether the participants completed PTT or the DCE first, *Order*. Because each participant saw a single level of benefit in each threshold series (a 40-percentage point reduction [from 60% to 20%] in the chance of developing RA), the relationship between the level of benefit and the risk threshold was estimated cross-sectionally.

The base interval regression was specified as:

$$Threshold_r = \alpha_r + \beta_1 Order_i + \varepsilon_r$$

where the intercept α_r represents the MAR for risk r for the 40-percentage point reduction in the chance of developing RA and ε_{br} is an independent and identically normally distributed random error term with a mean of zero and variance σ^2.

2.8.4 Heterogeneity in MAR

An interval regression was conducted for the PTT data, including clinically relevant covariates (age, education, health literacy, subjective numeracy, reported family history of RA) to test whether they were associated with MAR values. For each covariate that was significantly associated with the MAR in the interval regression for the PTT, separate subgroup analyses were conducted on the DCE data using the RPL modeling procedure described above.

2.8.5 Sensitivity Analyses

Several differences exist in how data from a DCE and PTT are modeled. To explore whether differences in the results appeared to be due to fundamental differences in the modeling approach or whether the differences in results likely can be explained by assumptions used in the applications of the methods in this study, several sensitivity analyses were conducted. These analyses should be interpreted with caution as these are post-hoc analyses meant to enhance comparison of

the results of the two methods by controlling for differences in the applications of the methods. By doing so, we make the analyses less in line with the formal recommendations on applying the (particularly DCE) analyses [14, 21]. The following sensitivity analyses were conducted:

1. In the primary interval regression for the PTT, the upper bound for the interval of greatest magnitude of each risk was set at the value of the MAR expressed by the respondent if the respondent indicated that they would accept the maximum level of risk included in the PTT questions. A separate interval regression was conducted in which the upper bound of this interval was set at 100%.
2. In the primary RPL model for the DCE, an interaction between 'Chance of developing RA' and 'Risk of serious infection' was included. A separate RPL model was estimated for the DCE data in which this interaction was excluded so that the DCE and PTT specifications were similar.
3. In the primary RPL model for the DCE, risk was modeled as categorical and the disutility of risk was allowed to be nonlinear. Because the benefit in the PTT analyses is fixed, there is no need to assume linearity or nonlinearity in the disutility of risk. To determine if assumptions about the linearity of the disutility of risk impacted the comparison between the DCE and PTT results, a separate RPL model was estimated for the DCE data in which all risk attributes were assumed to be linear.
4. A latent class analysis model was used to analyze the DCE data. Each covariate that was significantly associated with the MAR in the interval regression for the PTT was included in the class membership probability function in this analysis to test whether the covariates that explained differences in risk tolerance in the PTT data also explained differences in risk tolerance in the DCE data.

3 Results

The survey data of 982 (332 male, 650 female) individuals were included in the analyses. A full overview of the demographics including age, education, and health literacy can be found in Table 2.

3.1 Response to Feedback Questions

The majority of the participants reported that the choice tasks of the DCE and PTT were easy or very easy both to understand (71.7% and 64.7%, respectively) and to answer (59.8% and 55.6%, respectively) [Table 2]. However, there was a statistically significant ($p < 0.05$) 7-percentage point difference in the number of respondents indicating that the exercise was easy or very easy to understand between the DCE and the PTT, with more respondents indicating that the DCE was (very) easy to understand.

3.2 Avoiding Risks for DCE and PTT

Based on responses to the DCE, 4.8% of participants ($n = 47$) always chose to opt out (i.e., would avoid any risk of side effects by not taking a treatment). Similarly, for the PTT, some participants indicated they were not willing to accept any chance of getting mild side effects (5.2%), serious infection (6.1%), or serious side effects (4.6%). In the PTT, in total, 3.1% of participants ($n = 27$) indicated that they were unwilling to take any risks. Of these 27 participants unwilling to take any risk in the PTT, 12 (44%) always opted out in the DCE as well.

3.3 DCE Outcomes: Preferences

Table 3 shows the outcomes of the DCE analysis (using the RPL model), which were in line with a priori expectations. Participants preferred treatment with increasing effectiveness; they preferred a pill over an injection or drip, lower frequencies of administration over higher frequencies of administration, and lower chances of mild side effects, serious infections, and serious side effects over higher chances of getting these side effects and infections.

3.4 PTT Outcomes: Frequencies

Figure 1 shows the choice frequency for MAR intervals of risk attributes for a 40-percentage point decrease in the chance of developing RA. Some participants were willing to accept a 100% chance of mild side effects (2.2%) or serious infection (0.9%), while none of the participants was willing to accept a 100% chance of serious side effects. Most participants were willing to accept a mild side effect risk between 10 and 30%, a serious infection risk between 5 and 20%, and a serious side effect risk between 0.1 and 1% (42.5%, 42.6%, and 52.6%, respectively).

3.5 MAR

Table 4 shows the MAR calculated based on aggregate means estimates from the DCE and individual-level estimates from the DCE and PTT. For a 40-percentage point reduction in the chance of developing RA in the upcoming 2 years, participants were willing to accept a 50.8-percentage point increase in the chance of getting mild side

Table 2 Demographic variables and survey feedback responses for the total sample ($N = 982$)

	Frequency %
Age, years	
18–29	22.4
30–39	21.6
40–49	23.3
50–59	20.7
60–69	8.6
70 and over	3.5
Highest level of education	
None	1.2
Primary school	0.2
Secondary school	21.1
Sixth form	27.6
Degree or vocational	38.0
Postgraduate	11.1
Other	0.8
Family history of RA[2]	
Definitely not	21.4
Probably not	20.9
Don't know	24.9
Probably	16.6
Definitely	16.2
Smoking	
Yes	22.2
Never	54.6
No, but have in the past	23.2
Mean (SD)	**Median**
SILS (5-point scale; higher score, higher health literacy)	
4.27 (1.17)	5.00
SNS-3 total (higher score indicates higher numeracy)	
13.10 (3.72)	14.0
Perceived likelihood developing RA in the future	
3.14 (0.90)	3.00
	Frequency %
Understanding DCE	
Very easy to easy	71.7
Not easy to not difficult	20.7
Difficult to very difficult	7.6
Understanding PTT	
Very easy to easy	64.7
Not easy to not difficult	23.4
Difficult to very difficult	11.9
Answering DCE	
Very easy to easy	59.8
Not easy to not difficult	27.0
Difficult to very difficult	13.2
Answering PTT	
Very easy to easy	55.6
Not easy to not difficult	27.5
Difficult to very difficult	16.9

DCE discrete choice experiment, *PTT* probabilistic threshold technique, *RA* rheumatoid arthritis, *SD* standard deviation, *SILS* Single Item Literacy Screener, *SNS-3* 3-Item Version of the Subjective Numeracy Scale

effects, a 14.0-percentage point increase in the chance of getting a serious infection, and a 1.2-percentage point increase in the chance of serious side effects based on a DCE analysis using aggregate mean estimates. Based on the individual-level estimates, MAR values were comparable to overlapping confidence intervals. Based on the PTT analysis, these values were 18.1 percentage point, 10.2 percentage point, and 1.4 percentage point, respectively. Confidence intervals for the MAR based on the DCE and PTT overlapped for serious infection and serious side effects but not for mild side effects.

3.6 Heterogeneity in MAR

Significant preference heterogeneity was found based on health literacy, numeracy, and family history in the PTT analysis (Table 5). When conducting a subgroup analysis using RPL models for the DCE data, no significant differences were found as all 95% confidence intervals between subgroups overlapped (Table 5).

3.7 Sensitivity Analyses

Changes in the modeling of both the PTT and DCE resulted in slight changes in the numeric values for MAR (see ESM for the complete model output). Confidence intervals of the MAR estimates still overlapped for serious side effects and still did not overlap for mild side effects. Results related to the MAR estimates for serious infections were inconsistent. In the DCE model that excluded interactions or assumed the disutility of risk to be linear, the confidence intervals on the MAR estimates did not overlap with the confidence intervals of the MAR calculated based on the original PTT analysis. In contrast, the confidence intervals on the MAR estimated for serious infections for the original DCE overlap with the confidence intervals of the MAR estimates from the 'bounded' PTT analysis. Application of a latent class analysis model to identify heterogeneity in preferences and MAR estimates resulted in a five-class model (see ESM for the complete model output) with a class assignment model in which only subjective numeracy significantly impacted preferences and thus subsequently MAR. This is in contrast to the original PTT model in which health literacy and family history also impacted MAR estimates.

4 Discussion

This is the first paper that compares MAR estimates derived from DCE and PTT exercises. Findings showed that confidence intervals surrounding the MAR estimates

from the two methods overlapped for two out of the three risk attributes, namely 'chance to develop serious infection' and 'chance to develop serious side effects'. However, for the attribute 'chance to develop mild side effects', the MAR calculated based on the DCE data was significantly higher compared with the value calculated based on the PTT data.

The functional form of the attributes included in the MAR calculations likely influenced the outcomes. González Sepúlveda and Boeri showed that MAR outcomes were higher (though not statistically significantly so) when applying linear as compared with categorical specifications of the attributes used for MAR calculations in a DCE [34]. In the current study, the DCE-based MAR calculation for two out of the three attributes assumed that the levels were categorical, and that utility was linear between the levels. However, only for one attribute this resulted in a substantial and significant difference with the outcomes of the PTT (where linearity was assumed).

In the current study, the MAR calculated for the PTT may significantly underestimate actual preferences because (1) linearity was assumed in the PTT-based MAR calculation and (2) all key stakeholders (i.e., representatives from pharmaceutical industry, clinicians, and patients [24, 25]) included as research partners determined the appropriateness of the assumed functional forms in the DCE. The underestimation of the PTT-based MAR compared with the DCE-based MAR might be partially explained by the multi-attribute approach of a DCE, where respondents are forced to trade-off multiple attributes at the time, which results in the disclosure of the importance of each attribute relative compared to all the other attributes included in the experiment [14].

In the PTT, all risks were investigated separately, thereby not constraining the importance of a single attribute relative to all other attributes, potentially resulting in lower MAR for the 'less important' risks such as the chance of mild side effects. Additionally, the responses to the PTT might have been subject to anchoring effects [16]. Although respondents were able to indicate they would accept a relatively high risk or a certainty of side effects (i.e., 100% risk of side effects), their answers to the choice questions in the PTT were likely influenced by the level of risk provided in the initial choice tasks (although the risk level used was based on clinical insight into the likely risk of mild side effects and the range of risk levels was equal to that in the DCE). If the actual choice decision has a natural starting point that is known with certainty, then any anchoring effect in the study may simply reflect reality. However, when the starting point is known but uncertain or unknown, anchoring effects might introduce bias in results.

Future studies should explore: (1) if and to what extent results derived using the PTT are subject to anchoring effects (2); whether this might be influenced by the importance of the risk at stake (e.g., whether this effect is more

Table 3 Random parameter logit model output based on discrete choice experiment data

		Mean			SD		
Attribute	Level	Estimate	SE	95% CI	SD	SE	95% CI
No treatment (opt-out)		−3.60***	0.18	−3.95, −3.26	4.09***	0.22	3.66, 4.51
Chance of developing RA		−0.05***	0.00	−0.06, −0.04	0.06***	0.00	0.05, 0.07
Mode of administration	Tablet (reference)	0.56***	0.05	0.46, 0.66			
	Injection	−0.09***	0.03	−0.15, −0.04	0.54***	0.03	0.47, 0.60
	Infusion	−0.47***	0.04	−0.55, −0.39	0.81***	0.04	0.73, 0.89
Frequency of administration		0.21***	0.03	0.17, 0.26	0.35***	0.03	0.29, 0.42
Chance of mild side effects	2% (ref)	0.10***	0.02	0.05, 0.14			
	5%	0.06***	0.02	0.02, 0.09	0.15***	0.04	0.07, 0.23
	10%	−0.15***	0.02	−0.20, −0.10	0.27***	0.03	0.20, 0.33
Chance of serious infection		−0.14***	0.02	−0.18, −0.11	0.15***	0.01	0.13, 0.18
Chance of serious side effects	1 in 100,000 (ref)	0.33***	0.03	0.27, 0.40			
	20 in 100,000	−0.11***	0.02	−0.15, −0.06	0.08	0.12	−0.16, 0.32
	100 in 100,000	−0.23***	0.03	−0.29, −0.16	0.61***	0.03	0.54, 0.68
Chance of developing RA × risk of serious infection		0.002***	0.00	0.00;0.00	-	-	-
Log likelihood		−11713.7					
AIC		23469.5					
Pseudo R^2		0.28					

AIC Akaike information criterion, *CI* confidence interval, *RA* rheumatoid arthritis, *SD* standard deviation, *SE* standard error, ***$p < 0.001$

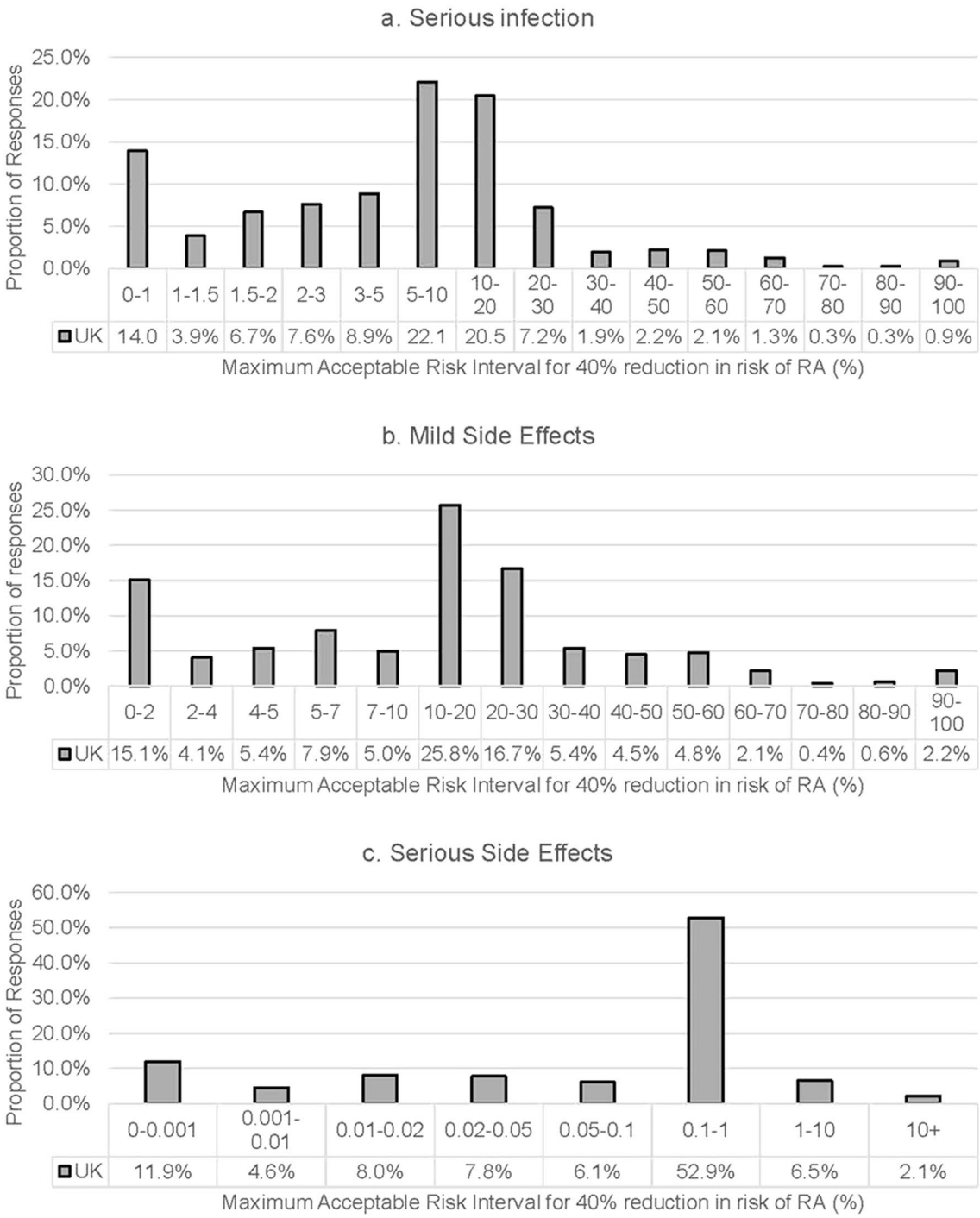

Fig. 1 Frequencies of thresholds of maximum acceptable risk intervals for a 40-percentage point reduction in the chance of developing rheumatoid arthritis (RA) in the upcoming 2 years based on a probabilistic threshold technique analysis, separately shown for (**A**) serious infection, (**B**) mild side effects, and (**C**) serious side effects

pronounced for relatively less important risk attributes); and (3) to what extent anchoring effects might be acceptable in cases where the risk levels included in the initial choice task are informed by clinical evidence for which there are different levels of certainty.

A potential advantage of the PTT relative to the DCE is the ability to identify heterogeneity in MAR. A number of

Table 4 MAR values calculated based on DCE average estimates, individual estimates, and PTT

	DCE: mean estimates		DCE: individual estimates		PTT	
	Mean (SE)	95% CI	Mean (SE)	95% CI	Mean (SE)	95% CI
MAR: 40% point RA risk reduction						
Chance of mild side effects[a]	50.8 (8.2)	34.7;66.8	42.6 (23.1)	−2.7;87.8	18.1 (0.9)	16.3;19.9
Chance of serious infection	14.0 (1.6)	10.9;17.1	20.6 (25.2)	−28.9;70.1	10.2 (0.7)	8.9;11.6
Chance of serious side effects[a]	1.2 (0.4)	0.5;2.0	2.2 (1.2)	−0.2;4.5	1.4 (0.3)	0.8; 2.0

CI confidence interval, *DCE* discrete choice experiment, *MAR* maximum acceptable risk, *PTT* probabilistic threshold technique, *SE* standard error

[a]These attributes were effects coded, linearity was assumed between levels and extrapolated beyond the levels included in the DCE for the purpose of the MAR calculation

Table 5 MAR values from the PTT based on interval regression including relevant covariates and MAR values from the DCE based on a subgroup analysis for significant covariates from the PTT interval regression

		Chance of mild side effects		Chance of serious infection		Chance of serious side effects	
		Mean (SE)	95% CI	Mean (SE)	95% CI	Mean (SE)	95% CI
PTT estimates							
Mean MAR (intercept)		17.1 (1.3)***		12.7 (0.98)***		2.1 (0.5)***	
Age	Young	1.0 (1.0)		1.2 (0.8)		0.2 (0.4)	
	Old	−1.0 (1.0)		−1.2 (0.8)		−0.2 (0.4)	
Education	Low	0.5 (0.6)		0.3 (0.5)		−0.3 (0.2)	
	High	−0.5 (0.6)		−0.3 (0.5)		0.3 (0.2)	
Health literacy	Low	0.0 (0.8)		2.6 (0.6)***		0.9 (0.3)***	
	High	−0.0 (0.8)		−2.6 (0.6)***		−0.9 (0.3)**	
Health numeracy	Low	−4.0 (0.7)***		−0.2 (0.5)		−0.2 (0.3)	
	High	4.0 (0.7)***		0.2 (0.5)		0.2 (0.3)	
Family history of RA	Not definitely	−1.9 (0.9)*		−2.7 (0.7)***		−1.1 (0.3)***	
	Definitely	1.9 (0.9)*		2.7 (0.7)***		1.1 (0.3)***	
DCE estimates							
Health literacy	Low	-[a]	-[a]	17.8 (8.7)	0.8;34.9	1.5 (0.9)	−0.2;3.2
	High	41.4 (6.4)	28.9;53.8	13.9 (1.5)	10.9;16.9	1.1 (0.4)	0.4;1.8
Numeracy	Low	47.7 (12.0)	24.1;71.3	6.1 (1.5)	3.2;9.1	-[a]	-[a]
	High	59.1 (14.4)	30.9;87.2	15.8 (2.1)	11.8;19.9	0.8 (0.2)	0.5;1.1
Family history	Definitely	30.4 (6.2)	18.3;42.5	14.6 (4.3)	6.3;22.9	1.0 (0.6)	−0.2;2.2
	Not definitely	47.5 (11.7)	34.5;80.5	14.1 (1.6)	10.9;17.3	1.0 (0.3)	0.5;1.6

CI confidence interval, *DCE* discrete choice experiment, *MA* maximum acceptable risk, *PTT* probabilistic threshold technique, *RA* rheumatoid arthritis, *SE* standard error, *$p < 0.1$; **$p < 0.05$; ***$p < 0.001$

[a]Sample size too small for a subgroup analysis

covariates that significantly impact MAR were identified using the interval regression on the PTT data (age, education, health literacy, numeracy, family history of RA), but these could not be identified using a subgroup RPL analysis on the DCE data. This shows the advantage of the PTT having MAR as a primary outcome measure calculated based on individual-level data. At the same time, this demonstrates the relative inefficiency of a sub-group analysis in DCE studies, even in relatively large samples such as in this case study. Although DCE studies have alternative strategies to investigate preference heterogeneity, for example, latent class analysis [14], such methods generally require large sample sizes.

More respondents indicated that the DCE choice tasks were easy or very easy to understand compared with the number of respondents who rated the PTT choice tasks the same way (a 7-percentage point difference that was statistically significant). Previous studies among similar populations without cognitive limitations reported similar findings [35–39]. This raises the question of whether

DCE choice tasks really are as difficult as researchers and clinicians have previously assumed; respondents' capabilities to accurately complete such questions might have been underestimated, which would call for a partial change in perspective toward this method as being (too) complex and time consuming to complete. However, it is essential to remember that this study used evidence-based guidelines for developing the DCE [19], whereas such guidelines do not yet exist for PTT studies. It is therefore possible that researchers in health economics are generally better trained in accurately designing and conducting DCE studies compared with PTT studies, as well as effectively incorporating the preferences of respondents related to the layout and presentation of information and choice tasks. With applications of PTT methodology being relatively novel and unexplored, further investigation into how best to design such experiments, when the aim is to measure MAR for multiple risks within one experiment, is warranted.

This study has considerable strengths, including a large sample size and the involvement of international multidisciplinary experts. However, this study also has some limitations. First, this is the first study to compare these methods. Therefore, definitive recommendations on the use of DCE versus PTT cannot be drawn from this study alone. However, although this analysis focuses on a sample of data collected in the UK, the study was also carried out in Germany and Romania [24], resulting in similar findings with respect to respondent feedback and MAR differences (see ESM), which supports the theoretical validity of this study. Second, this study was conducted among a sample of the general population who were members of a survey panel and did not consist of respondents with impaired cognitive skills. In addition, the current sample consisted of a relatively large proportion of respondents with a high educational level and/or a high health literacy level. Therefore, the current findings cannot be generalized across all (patient) populations. Third, several assumptions had to be made for MAR estimations. These assumptions might to some extent impact on the outcomes generated. This is especially true for the functional from of the attributes included [34]. Further studies should explore the impact of study assumptions related to method selection, functional form of attributes, and the data analysis strategy on MAR study outcomes and MPLC decision making. Fourth, to determine the maximum risk participants were willing to accept within the PTT method, an open-ended question was added to the survey in case participants indicated that they were willing to accept the highest presented risks in the preceding threshold questions. This procedure was used to limit the number of threshold questions in the survey. However, the literature suggests that responses to open-ended questions might provide a different threshold than one derived from bounded choice questions (e.g., [40, 41]). Because respondents whose MAR was greater than the highest level provided in the PTT exercise were allowed to state their MAR, this might have impacted the study results. Future research should be conducted to determine the best balance between estimating exact thresholds and determining upper bounds for acceptable risk levels in the PTT. Fifth, because a priori sample size calculations for DCE and PTT are complex, no formal power calculations were conducted. As a standard DCE includes about 100–300 respondents and a standard PTT include <100 respondents, this study oversampled and recruited 1000 respondents to ensure the responses provide enough information to identify preferences and allow comparisons across methods with acceptable precision.

5 Conclusions

Maximum acceptable risk estimates based on a DCE and a PTT were numerically similar with overlapping confidence intervals for two out of the three included risk attributes. With a 7-percentage point difference, the DCE was considered easier. This may suggest that the DCE is better suited in studies estimating MAR for multiple risk attributes of differing severity, while the PTT may be better suited when measuring heterogeneity in MAR estimates or when investigating one or more serious adverse events. Further studies should explore if this conclusion holds in different disease areas and respondent samples to provide further evidence on method selection guidance for estimating MAR to support stakeholder decision making on medical products.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40271-023-00643-w.

Acknowledgements This article and its contents reflect the view of the authors and not the view of PREFER, IMI, the European Union, or EFPIA.

Declarations

Funding This study formed part of the Patient Preferences in Benefit-Risk Assessments during the Drug Life Cycle (PREFER) project. The PREFER project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 115966. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. Karim Raza is supported by the National Institute for Health and Care Research Birmingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of Interest Rachael DiSantostefano and Ellen Janssen are employed by Janssen Research and Development, LLC and are stockholders in Johnson & Johnson. Brett Hauber is employed by Pfizer, Inc. and holds stock or stock options in Pfizer. Jorien Veldwijk, Gwenda Simons, Karin Schölin Bywall, and Marie Falahee have no conflicts of interest that are directly relevant to the content of this article. Karim Raza declared personal fees from Abbvie and Sanofi, and grant/research support from Bristol Myers Squibb. Matthias Englbrecht declares personal fees from AbbVie, Chugai, Eli Lilly, Novartis, Roche, Sanofi, and Mundipharma and grant/research support from Roche and Chugai.

Ethics Approval This study was approved by the London-Hampstead Research Ethics Committee (19/LO/0407).

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material Data are available upon reasonable request to the authors.

Code Availability Not applicable.

Authors' Contributions JV, RLD, and KR acquired funding. All authors were included in the final selection of methods for the study. GS, ME, KR, and MF led the clinical elements of the study including the attribute selection and contact with the patient research partner(s). JV, RLD, EJ, KSB, CR, and BH led the design for both methods. JV, GS, MF, and KR led survey programming and respondent recruitment. JV, EJ, and BH conducted the analyses. JV, RLD, EJ, and MF drafted the manuscript. All authors reviewed and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

References

1. de Bekker-Grob EW, Berlin C, Levitan B, Raza K, Christoforidi K, Cleemput I, et al. Giving patients' preferences a voice in medical treatment life cycle: the PREFER public–private project. Patient. 2017;10(3):263–6.
2. Ho M, Saha A, McCleary KK, Levitan B, Christopher S, Zandlo K, et al. A framework for incorporating patient preferences regarding benefits and risks into regulatory assessment of medical technologies. Value Health. 2016;19(6):746–50.
3. Marsh K, van Til JA, Molsen-David E, Juhnke C, Hawken N, Oehrlein EM, et al. Health preference research in Europe: a review of its use in marketing authorization, reimbursement, and pricing decisions: report of the ISPOR Stated Preference Research Special Interest Group. Value Health. 2020;23(7):831–41.
4. Brett Hauber A, Fairchild AO, Reed JF. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy. 2013;11(4):319–29.
5. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice xperiments in health economics: past, present and future. Pharmacoeconomics. 2019;37(2):201–26.
6. de Bekker-Grob EW, Swait JD, Kassahun HT, Bliemer MCJ, Jonker MF, Veldwijk J, et al. Are healthcare choices predictable? The impact of discrete choice experiment designs and models. Value Health. 2019;22(9):1050–62.
7. De Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait J. Can healthcare choice be predicted using stated preference data? Soc Sci Med. 2020;2:2.
8. Lambooij MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YW, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Med Res Methodol. 2015;15:19.
9. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ. 2018;19(8):1053–66.
10. Salampessy BH, Veldwijk J, JantineSchuit A, van den Brekel-Dijkstra K, Neslo RE, de Ardine WG, et al. The predictive value of discrete choice experiments in public health: an exploratory application. Patient. 2015;8(6):521–9.
11. Watson V, Becker F, de Bekker-Grob E. Discrete choice experiment response rates: a meta-analysis. Health Econ. 2017;26(6):810–7.
12. DeShazo JR, Fermob G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. J Environ Econ Manage. 2011;44(1):123–43.
13. Rose JM, Bliemer MJC. Constructing efficient stated choice experimental desings. Transport Rev. 2009;9:1–31.
14. Hensher D, Rose JM, Greene WH. Applied choice analysis. 2nd ed. Cambridge: Cambridge University Press; 2015.
15. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. Patient. 2015;8(5):373–84.
16. Hauber B, Coulter J. Using the threshold technique to elicit patient preferences: an introduction to the method and an overview of existing empirical applications. Appl Health Econ Health Policy. 2020;18(1):31–46.
17. Medical Device Innovation Consortium. Using the MDIC Patient Centered Benefit-Risk Framework to support an expanded indication. 2019. Available from: https://mdic.org/wp-content/uploads/2019/04/NxStage-Case-Study.pdf. [Accessed 30 Apr 2023].
18. US Food and Drug Administration. Summary of safety and effectiveness data (SSED): lidocaine/epinephrine iontophoresis and automated tympanostomy tube insertion system. 2019. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf19/P190016B.pdf. [Accessed 30 Apr 2023].
19. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health–a checklist: a report of the ISPOR good research practices for conjoint analysis Task force. Value Health. 2011;14(4):403–13. https://doi.org/10.1016/j.jval.2010.11.013.
20. Johnson FR, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health. 2013;16(1):3–13. https://doi.org/10.1016/j.jval.2012.08.2223.
21. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value Health. 2016;19(4):300–15.

22. Louviere JJ, Hensher DA, Swait JD. Stated choice methods; analysis and application. Cambridge: Cambridge University Press; 2000.
23. Hauber B, Mange B, Zhou M, Chaudhuri S, Benz HL, Caldwell B, et al. Parkinson's patients' tolerance for risk and willingness to wait for potential benefits of novel neurostimulation devices: a patient-centered threshold technique study. MDM Policy Pract. 2021;6(1):2381468320978407.
24. Falahee M, Simons G, DiSantostefano RL, Valor Méndez L, Radawski C, Englbrecht M, et al. Treatment preferences for preventive interventions for rheumatoid arthritis: protocol of a mixed methods case study for the Innovative Medicines Initiative PREFER project. BMJ Open. 2021;11: e045851.
25. Simons G, Veldwijk J, DiSantostefano RL, Englbrecht M, Radawski C, Bywall KS, et al. Preferences for preventive treatments for rheumatoid arthritis: discrete choice survey in the UK Germany and Romania. Rheumatology (Oxford). 2023;62(2):596–605.
26. Simons G, Janssen EM, Veldwijk J, DiSantostefano RL, Englbrecht M, Radawski C, et al. Acceptable risks of treatments to prevent rheumatoid arthritis among first-degree relatives: demographic and psychological predictors of risk tolerance. RMD Open. 2022;8(2): e002593.
27. Simons G, Caplan J, DiSantostefano RL, Veldwijk J, Englbrecht M, SchölinBywall K, et al. A systematic review of quantitative preference studies of treatments for rheumatoid arthritis among patients and at risk populations: attributes and attribute selection. Arthritis Res Ther. 2022;2:2.
28. Rose JM, Collins AT, Bliemer MCJ, Hensher DA. ChoiceMetrics. Ngene 1.1.1. 2012, user manual & reference guide. ChoiceMetrics, Sydney, Australia.
29. Morris NS, MacLean CD, Chew LD, Littenberg B. The single item literacy screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract. 2006;7:21.
30. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a short, 3-item version of the subjective numeracy scale. Med Decis Making. 2015;35(8):932–6.
31. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ. 2005;14(10):1079–83.
32. Hensher DA, Rose JM, Greene WH. Applied choice analysis: a primer. New York: Cambridge University Press; 2000.
33. Cameron AC, Trivedi PK. Microeconometrics using Stata. College Station: Stata Press; 2010.
34. Gonzalez JM, Boeri M. The impact of the risk functional form assumptions on maximum acceptable risk measures. Patient. 2021;14(6):827–36.
35. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van MM. COVID-19 contact tracing Apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth. 2020;8(10): e20741.
36. Veldwijk J, Johansson JV, Donkers B, de Bekker-Grob EW. Mimicking real-life decision making in health: allowing respondents time to think in a discrete choice experiment. Value Health. 2020;23(7):945–52.
37. Visser LA, Huls SPI, Uyl-de Groot CA, de Bekker-Grob EW, Redekop WK. An implantable device to treat multiple sclerosis: a discrete choice experiment on patient preferences in three European countries. J Neurol Sci. 2021;428: 117587.
38. Jonker MF, Donkers B, Goossens LMA, Hoefman RJ, Jabbarian LJ, de Bekker-Grob EW, et al. Summarizing patient preferences for the competitive landscape of multiple sclerosis treatment options. Med Decis Making. 2020;40(2):198–211.
39. Rutten-van Molken M, Karimi M, Leijten F, Hoedemakers M, Looman W, Islam K, et al. SELFIE Consortium. Comparing patients' and other stakeholders' preferences for outcomes of integrated care for multimorbidity: a discrete choice experiment in eight European countries. BMJ Open. 2020;10(10): e037547.
40. Bijlenga D, Bonsel GJ, Birnie E. Eliciting willingness to pay in obstetrics: comparing a direct and an indirect valuation method for complex health outcomes. Health Econ. 2011;20(11):1392–406.
41. van der Marjon P, Shiell A, Au F, Johnston D, Tough S. Convergent validity between a discrete choice experiment and a direct, open-ended method: comparison of preferred attribute levels and willingness to pay estimates. Soc Sci Med. 2008;67(12):2043–50.

Authors and Affiliations

Jorien Veldwijk[1,2,3] · Rachael Lynn DiSantostefano[4] · Ellen Janssen[4] · Gwenda Simons[5] · Matthias Englbrecht[6,7] · Karin Schölin Bywall[8] · Christine Radawski[9] · Karim Raza[5,10,11] · Brett Hauber[12,13] · Marie Falahee[5]

✉ Jorien Veldwijk
veldwijk@eshpm.eur.nl

1 School of Health Policy and Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

2 Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, The Netherlands

3 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

4 Janssen Research & Development, Titusville, NJ, USA

5 Rheumatology Research Group, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK

6 Freelance Healthcare Data Scientist, Greven, Germany

7 Department of Internal Medicine and Institute for Clinical Immunology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

8 Centre for Research Ethics & Bioethics, Uppsala University, Uppsala, Sweden

9 Eli Lilly and Company, Indianapolis, IN, USA

10 Department of Rheumatology, Sandwell and West Birmingham NHS Trust, Birmingham, UK

11 MRC Versus Arthritis Centre for Musculoskeletal Ageing Research and Research into Inflammatory Arthritis Centre Versus Arthritis, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

12 Pfizer, Inc., New York, NY, USA

13 The Comparative Health Outcomes, Policy and Economics (CHOICE) Institute, University of Washington School or Pharmacy, Seattle, WA, USA