CrossMark

PRACTICAL APPLICATION

Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide

Esther W. de Bekker-Grob · Bas Donkers ·
Marcel F. Jonker · Elly A. Stolk

Abstract Discrete-choice experiments (DCEs) have become a commonly used instrument in health economics and patient-preference analysis, addressing a wide range of policy questions. An important question when setting up a DCE is the size of the sample needed to answer the research question of interest. Although theory exists as to the calculation of sample size requirements for stated choice data, it does not address the issue of minimum sample size requirements in terms of the statistical power of hypothesis tests on the estimated coefficients. The purpose of this paper is threefold: (1) to provide insight into whether and how researchers have dealt with sample size calculations for healthcare-related DCE studies; (2) to introduce and explain the required sample size for parameter estimates in DCEs; and (3) to provide a step-by-step guide for the calculation of the minimum sample size requirements for DCEs in health care.

E. W. de Bekker-Grob and B. Donkers contributed equally to this work.

E. W. de Bekker-Grob (✉)
Department of Public Health, Erasmus MC, University Medical Centre Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands
e-mail: e.debekker@erasmusmc.nl

B. Donkers
Department of Business Economics, Erasmus University, Rotterdam, The Netherlands

M. F. Jonker · E. A. Stolk
Department of Health Economics, Policy and Law, Erasmus University, Rotterdam, The Netherlands

Key Points for Decision Makers

The minimum sample size needed for a discrete-choice experiment (DCE) depends on the specific hypotheses to be tested.

DCE practitioners should realize that a small size effect may still be meaningful, but that a limited sample size prevents detection of such small effects.

Policy makers should not make a decision on non-significant outcomes without considering whether the study had a reasonable power to detect the anticipated outcome.

1 Introduction

Discrete-choice experiments (DCEs) have become a commonly used instrument in health economics and patient-preference analysis, addressing a wide range of policy questions [1, 2]. DCEs allow for a quantitative elicitation of individuals' preferences for health care interventions, services, or policies. The DCE approach combines consumer theory [3], random utility theory [4], experimental design theory [5], and econometric analysis [1]. See Louviere et al. [6], Hensher et al. [7], Rose and Bliemer [8], Lancsar and Louviere [9], and Ryan et al. [10] for further details on conducting a DCE.

DCE-based research in health care is often concerned about establishing the impact of certain healthcare interventions and aspects (i.e., attributes) thereof on patients' decisions [11–20]. Consequently, a typical research

question is to establish whether or not individuals are indifferent between two attribute levels. For instance: Do patients prefer delivery at home more than in a hospital?; Do patients prefer a medical specialist over an nurse practitioner?; Do patients prefer every 5 year screening over every 10 year screening?; Do patients prefer a weekly oral medication over a monthly injection?; Do patients prefer the explanation of their medical results through a face-to-face contact more than through a letter? As a result, an important design question is the size of the sample needed to answer such a research question. When considering the required sample size, DCE practitioners need to be confident that they have sufficient statistical power to detect a difference in preferences when this difference is sufficiently large. A practical solution (that does not require any sample size calculations) is to simply maximize the sample size given the research budget at hand, i.e., trying to overpower the study as much as possible. This is beneficial for reasons other than statistical precision (e.g. to facilitate in-depth analysis). However, particularly in the health care area, the number of eligible patients and healthcare professionals is generally limited. Although theory exists as to the calculation of sample size requirements for stated choice data, it does not address the issue of minimum sample size requirements in terms of testing for specific hypotheses based on the parameter estimates produced [21].

The purpose of this paper is threefold. The first objective is to provide insight into whether and how researchers have dealt with sample size calculations for health care-related DCE studies. The second objective is to introduce and explain the required sample size for parameter estimates in DCEs. The final objective of this manuscript is to provide a step-by-step guide for the calculation of the minimum sample size requirements for DCEs in healthcare.

2 Literature Review

2.1 Methods

To gain insight into the current approaches to sample size determination, we reviewed health care-related DCE studies published in 2012. Older literature was ignored, as the research frontier for methodological issues has shifted a lot over the past years [1, 22]. MEDLINE was used to identify healthcare-related DCE studies, replicating the methodology of two comprehensive reviews of the healthcare DCE literature [1, 2]. The following search terms were used: conjoint, conjoint analysis, conjoint measurement, conjoint studies, conjoint choice experiment, part-worth utilities, functional measurement, paired comparisons, pairwise choices, discrete choice experiment, dce,

discrete choice mode(l)ling, discrete choice conjoint experiment, and stated preference. Studies were included if they were choice-based, published as a full-text English language article, and applied to healthcare. Consideration was given to background information of the studies, and detailed consideration was given to whether and how sample size calculations were conducted. We also briefly describe the methods that have been used to obtain sample size estimates so far.

2.2 Literature Review Results

The search generated 505 possible references. After reading abstracts or full articles, 69 references met the inclusion criteria. The appendix shows the full list of references [Electronic Supplementary Material (ESM) 1]. Table 1 summarizes the review data. Most DCE studies were from the UK, with the USA, Canada, and Australia also major contributors. Studies having 4–6 attributes and 9–16 choice sets per respondent were commonly used in the published healthcare-related DCE studies in 2012. The sample sizes differed substantially between the DCE studies.

Of 69 DCEs, 22 (32 %) had sample sizes smaller than 100 respondents, whereas 16 (23 %) of the 69 DCEs had sample sizes larger than 600 respondents; six (9 %) DCEs even had sample sizes larger than 1000 respondents. More than 70 % of the DCE studies (49 of 69) did not (clearly) report whether and what kind of sample size method was used; 12 % of the studies (8 of 69) just referred to other DCE studies to explain the sample size used. For example, Huicho et al. [23] mentioned that "Based on the experience of previous studies [24, 25], we aimed for a sample size of 80 nurses and midwives", and Bridges et al. [26] mentioned "In a previously published pilot study, the conjoint analysis approach was shown to be both feasible and functional in a very low sample size (n = 20) [27]". In 13 % of the DCE studies (9 of 69 [28–36]), one or more of the following rules of thumb were used to estimate the minimum sample size required: that proposed by (1) Johnson and Orme [37, 38]; (2) Pearmain et al. [39]; and/or (3) Lancsar and Louviere [9].

In short, the rule of thumb as proposed by Johnson and Orme [37, 38] suggests that the sample size required for the main effects depends on the number of choice tasks (t), the number of alternatives (a), and the number of analysis cells (c) according to the following equation:

$$N > 500c/(t \times a) \tag{1}$$

When considering main effects, 'c' is equal to the largest number of levels for any of the attributes. When considering all two-way interactions, 'c' is equal to the largest product of levels of any two attributes [38].

Table 1 Background information and sample size (method) used of published health care-related discrete-choice experiment studies in 2012 (N = 69)

Item	N (%)
Country of origin[a]	
UK	16 (23)
USA	13 (19)
Canada	10 (14)
Australia	7 (10)
Germany	6 (9)
Netherlands	4 (6)
Denmark	3 (4)
Other	19 (28)
Number of attributes[a]	
2–3	5 (7)
4–5	24 (35)
6	25 (36)
7–9	17 (25)
>9	3 (4)
Number of choices per respondent	
8 or fewer	14 (20)
9–16 choices	47 (68)
More than 16 choices	5 (7)
Not clearly reported	3 (4)
Sample size used[a]	
<100	22 (32)
100–300	28 (41)
300–600	17 (25)
600–1,000	10 (14)
>1,000	6 (9)
Sample size method used[a]	
Parametric approach	4 (6)
Louviere et al. [6]	3 (4)
Rose and Bliemer [21]	1 (1)
Rule of thumb	9 (13)
Johnson and Orme [28, 29]	5 (7)
Pearmain et al. [30]	2 (3)
Lancsar and Louviere [9]	3 (4)
Referring to studies	8 (12)
Review studies	3 (4)
Applied studies	5 (7)
Not (clearly) reported	49 (71)

[a] Totals do not add up to 100 % as some studies were conducted in different countries, used a different number of attributes per discrete-choice experiment, used several subgroups of respondents, and/or used multiple sample size methods

The rule of thumb proposed by Pearmain et al. [39] suggests that, for DCE designs, sample sizes over 100 are able to provide a basis for modeling preference data, whereas Lancsar and Louviere [9] mentioned "our empirical experience is that one rarely requires more than 20 respondents per questionnaire version to estimate reliable models, but undertaking significant post hoc analysis to identify and estimate co-variate effects invariably requires larger sample size".

Four of 69 (6 %) reviewed DCE studies used a parametric approach to estimate the minimum sample size required (a parametric approach can be used if one assumes, for example based on the law of large numbers, that the focal quantity—an estimated probability or coefficient—is Normally distributed. This assumption facilitates the derivation of the minimum sample sizes required). That is, three studies used the parametric approach as proposed by Louviere et al. [6] and one study [40] reported the parametric approach as proposed by Rose and Bliemer [21]. Louviere et al. [6] assume the study is being conducted to measure a choice probability with some desired level of accuracy. The asymptotic sampling distribution (i.e., the distribution as sample size $N \to \infty$) of a proportion p_N, obtained by a random sample of size N, is Normal with mean p (the true population proportion) and variance pq/N, where $q = 1-\mathrm{p}$. The minimum sample size to estimate the true proportion within α_1 % of the true value p with a probability α_2 or greater has to satisfy the requirement that $\mathrm{Prob}(|p_N-p| \leq \alpha_1 p) \geq \alpha_2$, which can be calculated using the following equation:

$$N > (q/(rpa_1^2)) \cdot \left(\Phi^{-1}(1 - \frac{\alpha_2}{2}) \right)^2 \qquad (2)$$

where Φ^{-1} is the inverse cumulative Normal distribution function, and r is the number of choice sets per respondent. Hence, the parametric approach as proposed by Louviere et al. [6] suggests that the sample size required for the main effects depends on the number of choice sets per respondent (r), the true population proportion (p), the one minus true population proportion (q), the inverse cumulative Normal distribution function (Φ^{-1}), the allowed deviation from the true population proportion (α_1), and the significance level (α_2).

The parametric approach that has been recently introduced by Rose and Bliemer [21] focuses on the minimum sample size required based on the most critical parameter (i.e., to be able to determine whether each parameter value is statistically significant from zero). This parametric approach can only be used if prior parameter estimates are available and not equal to zero. The minimum required sample size to state with 95 % certainty that a parameter estimate is different from zero can be determined according to the following equation:

$$N > \max_k (1.96\sqrt{\sum_{\gamma k}}/\gamma k)^2 \qquad (3)$$

where γ_k is the parameter estimate of attribute k, and $\Sigma_{\gamma k}$ is the corresponding variance of the parameter estimate of attribute k.

2.3 Comment on the State of Play

The disadvantage of using one of the rules of thumb mentioned in paragraph 2.2 is that such rules are not intended to be strictly accurate or reliable. The parametric approach as proposed by Louviere et al. [6] is not suitable for determining the minimum required sample size for coefficients in DCEs, as this approach focuses on choice probabilities and does not address the issue of minimum sample size requirements in terms of testing for specific hypotheses based on the parameter estimates produced. The parametric approach for minimum sample size calculation proposed by Rose and Bliemer [21] is solely based on the most critical parameter, so it is not specific to a certain hypothesis. It also does not depend on a desired power level for the hypothesis tests of interest.

3 Determining Required Sample Sizes for Discrete-Choice Experiments (DCEs): Theory

In this section we explain the analysis needed to determine the minimum sample size requirements in terms of testing for specific hypotheses for coefficients in DCEs. Our proposed approach is more general than the parametric approaches mentioned in Sect. 2, as it can be used for any particular hypothesis that is relevant to the researcher. We outline which elements are required before such a minimum sample size can be determined, why these elements are needed, and how to calculate the required sample size. To provide a step-by-step guide that is useful for researchers from all different kinds of backgrounds, we strive to keep the number of formulas in this section as low as possible. Nevertheless, a comprehensive explanation of the minimum sample size calculation for coefficients in DCEs can be found in the appendix (ESM 2).

3.1 Required Elements for Estimating Minimum Sample Size

Before the minimum sample size for coefficients in a DCE can be calculated, the following five elements are needed:

- Significance level (α)
- Statistical power level ($1-\beta$)
- Statistical model used in the DCE analysis [e.g., multinomial logit (MNL) model, mixed logit (MIXL) model, generalized multinomial logit (G-MNL) model]
- Initial belief about the parameter values
- The DCE design.

3.1.1 Significance Level (α)

The significance level α sets the probability for an incorrect rejection of a true null hypothesis. For example, if one wants to be 95 % confident that the null hypothesis will not be rejected when it is true, α needs to be set at $1-0.95 = 0.05$ (i.e. 5 %). Conversely, if one decides to perform a hypothesis test at a $1-\alpha$ confidence level, there is by definition an α probability of finding a significant deviation when there is in fact no true effect. Perhaps unsurprisingly, the smaller the imposed value of α (i.e., the more certainty one requires), the larger the minimum required sample size will be.

3.1.2 Statistical Power Level ($1-\beta$)

β indicates the probability of failing to reject a null hypothesis when the null hypothesis is actually false. The chosen value of beta is related to the statistical power of a test (which is defined as $1-\beta$). As we want to assess whether a parameter value (coefficient) is significantly different from zero, we can define the sample size that enables us to find a significant deviation from zero in at least $(1-\beta) \times 100$ % of the cases. For example, a statistical power of 0.8 (or 80 %) means that a study (when conducted repeatedly over time) is likely to produce a statistically significant result eight times out of ten. A larger statistical power level will increase the minimum sample size needed.

3.1.3 Statistical Model Used in the DCE Analysis

The calculation of the minimum required sample size also depends on the type of statistical model that will be used to analyze the DCE data (e.g., MNL, MIXL, G-MNL). The type of statistical model affects the number of parameters that needs to be estimated, the corresponding parameter values, and the parameter interpretation. As a consequence, the estimation precision of the parameters, which we will characterize through the variance covariance matrix of the estimated parameters, also depends on the statistical model that is used. In order to properly determine the estimation precision of each of the parameters, the statistical model needs to be specified.

3.1.4 Initial Belief About the Parameter Values

Of course, if the true values of the parameters (coefficients) were known, one would not need to execute the DCE. Nevertheless, before a minimum sample size can be determined, an initial estimate of the parameter values is required for two reasons. First, in models that are nonlinear

in the parameters, such as choice models, the asymptotic variance–covariance matrix (AVC) depends on the values of the parameters themselves. This AVC is an intermediate stage in the sample size calculation (see Sect. 3.2 for more details), and reflects the expected accuracy of the statistical estimates obtained using the statistical model as identified under Sect. 3.1.3. Second, before a power calculation can be done, one has to describe a specific hypothesis and the power one wants to achieve given a certain degree of misspecification (i.e., the degree to which the true coefficient value deviates from its hypothesized value). As null hypothesis, we will use the hypothesis that there is no influence so the coefficient equals zero. The initial estimate of the parameter value can then be used as value for the effect size. The closer to zero the effect size is, the more difficult it will be to find a significant effect and hence the larger the minimum sample size will be. To obtain some insight into these parameter values, a small pilot DCE study—for example with 20–40 respondents—may be helpful.

3.1.5 DCE Design

The large literature on efficient design generation indicates the importance of the design in getting accurate estimates and powerful tests. The DCE design is described by the number of choice sets, the number of alternatives per choice set, the number of attributes, and the combination of the attribute levels in each choice set. The DCE design has a direct influence on the AVC, which affects the estimation precision of the parameters, and hence will have a direct influence on the minimum sample size required.[1]

3.2 Sample Size Calculation for DCEs

Once all five required elements mentioned in Sect. 3.1 have been determined, the minimum required sample size for the estimated coefficients in a DCE can be calculated. First, as an intermediate part of the sample size calculation, the AVC has to be established. That is, the statistical model (Sect. 3.1.3), the initial belief on the parameter values, denoted with γ (Sect. 3.1.4), and the DCE design (Sect. 3.1.5), are all needed to infer the AVC matrix, $\sum_\gamma$, of the estimated parameters. Details on how to construct the variance–covariance matrix from this information can be found, for example, in McFadden [4] for MNL and in

Bliemer and Rose [41] for panel MIXL. A variance–covariance matrix is a square matrix that contains the variances and covariances associated with all the estimated coefficients. The diagonal elements of this matrix contain the variances of the estimated coefficients, and the off-diagonal elements capture the covariances between all possible pairs of coefficients. For hypothesis tests on individual coefficients, we only need the diagonal elements of $\sum_\gamma$, which we denote by $\Sigma_{\gamma k}$ for the kth diagonal element.

Once the AVC, $\sum_\gamma$, of the estimated parameters has been established and the confidence level (α), the power level ($1-\beta$), and the effect sizes (δ) are set, the minimum required sample size (N) for the estimated coefficients in a DCE can be calculated (see Eq. 4).

$$N > \left((z_{1-\beta} + z_{1-\alpha}) \sqrt{\sum_{\gamma k}} / \delta \right)^2 \tag{4}$$

Each of the elements in this sample size calculation intuitively makes sense. In particular, with a larger effect size δ, a smaller sample size (N) will suffice to have enough power to find a significant deviation. Testing at a higher confidence level α increases $z_{1-\alpha}$,[2] and thus increases the minimum required sample size (N). The same holds when more statistical power is desired, as this increases $z_{1-\beta}$.[3] When the variance-covariance matrix contains smaller variance ($\sum_{\gamma k}$) the minimum sample size (N) required decreases, as the estimates will be more precise. Smaller values for $\sum_{\gamma k}$ can be obtained from using more choice sets, more alternatives per choice set or a more efficient design.

4 Determining Required Sample Sizes for DCEs: A Practical Example

In this section, a practical example is provided to explain, step-by-step, how the minimum sample size requirement

---

[1] All aspects of our sample size calculation are conditional on the design of the experiment and the implementation in a questionnaire. The survey design will have an impact on the precision of the parameters that should be accounted for through its effect on the anticipated parameter values. Also, the model specification has an impact on the precision of the parameters.

[2] The value of α (Sect. 3.1.1) is used to determine the corresponding quantile of the Normal distribution ($z_{1-\alpha}$) that is needed in the sample size calculations. The value of $z_{1-\alpha}$ for a given α can be found in the basic statistics textbooks or easily calculated in Microsoft Excel® using the formula NORMSINV($1-\alpha$). The value of $z_{1-\alpha}$ for an α of 0.05 equals 1.64.

[3] In the computation of the sample size, we need $z_{1-\beta}$, the quantile of the Normal distribution with $\Phi(z_{1-\beta}) = 1-\beta$. Here again, Φ denotes the cumulative distribution function of the Normal distribution. Accordingly, the value for $z_{1-\beta}$ for a given $1-\beta$ can be found in the basic statistics textbooks or easily calculated in Microsoft Excel® using the formula NORMSINV($1-\beta$); e.g., assuming a statistical power level of 80 %, the value $z_{1-\beta}$ is 0.84 [i.e., NORMSINV(0.8)].

Table 2 Alternatives, attributes and levels for preventive osteoporosis drug treatment, their parameter labels, initial belief about parameter values, and discrete-choice experiment design codes (based on de Bekker-Grob et al. [12])

		Parameter label	Initial belief parameter value	DCE design code
Alternative	Alternative label			
Constant (i.e., alternative specific constant for drug treatment; intercept)		A	1.23	
Alternative 1	Drug treatment alternative I			1
Alternative 2	Drug treatment alternative II			1
Alternative 3	Opt-out alternative			0
Attribute	Attribute levels			
Drug administration	Tablet once a month			
	Tablet once a week	B1	–0.31	1
	Injection every 4 months	B2	–0.21	1
	Injection once a month	B3	–0.44	1
Effectiveness (%)		C	0.028	
	5			5
	10			10
	25			25
	50			50
Side effect nausea		D	–1.10	
	No			0
	Yes			1
Treatment duration (years)		E	–0.04	
	1			1
	2			2
	5			5
	10			10
Cost (€)		F	–0.0015	
	0			0
	120			120
	240			240
	720			720

for a DCE study can be calculated. This is illustrated using R-code, which can also be found at http://www.erim.eur.nl/ecmc.

The DCE study used for this illustration concerns a DCE about patients' preferences for preventive osteoporosis drug treatment [12]. In this DCE study, patients had to choose between drug treatment alternatives that differed in five treatment attributes: route of drug administration, effectiveness, side effects (nausea), treatment duration, and out-of-pocket costs. The DCE design was orthogonal and contained 16 choice sets. Each choice set consisted of two unlabeled drug treatment alternatives and an opt-out option.

In what follows, we show in seven steps how the minimum sample size for coefficients can be calculated for the DCE on patients' preferences for preventive osteoporosis drug treatment.

Step 1   Significance Level (α)

We first have to set the confidence through α. In the illustration, we choose $\alpha = 0.05$. The resulting confidence level is 95 %, assuming a one-tailed test[4] (Box 1)

Box 1

```
test_alpha=0.05
z_one_minus_alpha<-qnorm(1-test_alpha)
```

---

[4] A one-tailed test is used if only deviations in one direction are considered possible; in contrast, a two-tailed test is used if deviations of the estimated parameter in either direction from zero are considered theoretically possible. Be aware that, for a two-tailed test, the alpha level should be divided by 2 (i.e., $\alpha/2$).

Step 2   Statistical Power Level $(1-\beta)$
The second step is to choose the statistical power level. For our illustration, we opt for a standard statistical power level of 80 % (i.e., $\beta = 0.20$, hence $1-\beta = 0.80$) (Box 2).

Box 2

```
test_beta=0.20
z_one_minus_beta<-qnorm(1-test_beta)
```

Step 3   Statistical Model Used in the DCE Analysis
The third step is to choose the statistical model to analyze the DCE data. For our illustration, we opt for an MNL model. In the R code, this affects the way the AVC needs to be calculated, which is outlined in step 6

Step 4   Initial Belief About the Parameter Values
The fourth step concerns the initial beliefs about the parameter values. The DCE illustration regarding patients' preferences for preventive osteoporosis drug treatment contains five attributes (two categorical attributes and three linear attributes) [12], resulting in eight parameters to be estimated (see Table 2 column 'parameter label'). We use the point estimates of the parameters as our guess of the coefficients and the effect sizes δ (see Table 2 column 'initial belief parameter value') (Box 3)

Box 3

```
parameters<-c(1.23 , -0.31 , -0.21 , -0.44 , 0.028 ,
-1.10 , -0.04 , -0.0015)
```

Step 5   The DCE design
The fifth step focuses on the DCE design. The DCE design requires eight parameters to be estimated (ncoefficients = 8). Each choice set contains three alternatives (nalts = 3); that is, two drug treatment alternatives, and one opt-out alternative. The DCE design contains 16 choice sets (nchoices = 16) (Box 4)

Box 4

```
ncoefficients=8
nalts=3
nchoices=16
```

The DCE design should be coded in a text-file in such a way that it can be read correctly into R. That is, the DCE design should contain one row for each alternative. So, there should be nalts × nchoices rows (see Table 3 as an example for our illustration, which contains 48 rows (i.e., 3 alternatives × 16 choice sets); rows 1–3 correspond to choice set 1, rows 4–6 correspond to choice set 2, etc.)

Each row should contain the coded attribute levels for that alternative. See Table 3 for how the DCE design for our illustration was coded (columns A–F). For example, row 1 corresponds to the first preventive drug treatment alternative in choice set 1: a drug treatment alternative (value 1, column A) that should be taken as a tablet every week (value 1, column B1), which will result in a 5 % reduction of a hip fracture (value 5, column C) without side effects (value 0, column D), for which the drug treatment duration will be 10 years (value 10, column E) and out-of-pocket costs of €120 are required (value 120, column F). Be aware that only the DCE design (i.e., the 'white part' of Table 3) should be in a text file, so that it can be read correctly in R (Box 5)

Box 5

```
# load the design information
design<-as.matrix(read.table("….fill in the file name
 of the DCE design…",header=FALSE));
```

Step 6   Estimation Accuracy
          Having our statistical model, our initial beliefs about the parameter values (i.e., our guess of the effect sizes) and our DCE design matrix, we are able to compute the AVC matrix ($\sum_{\gamma}$) (Box 6)

The results of the minimum sample size required to obtain the desired power level for finding an effect when testing at a specific confidence level for each parameter are shown in Table 4. To illustrate the impact of the

---

Box 6

```
#compute the information matrix, see Appendix (Electronic Supplementary Material 2) for more details
# initialize a matrix of size ncoefficients by ncoefficients filled with zeros.
info_mat=matrix(rep(0,ncoefficients* ncoefficients), ncoefficients, ncoefficients)
# compute exp(design matrix times initial parameter values)
exputilities=exp(design%*%parameters)
# loop over all choice sets
for (k_set in 1:nchoices) {
# select alternatives in the choice set
alternatives=((k_set-1)*nalts+1) : (k_set*nalts)
# obtain vector of choice shares within the choice set
p_set=exputilities[alternatives]/sum(exputilities[alternatives])
# also put these probabilities on the diagonal of a matrix that only contains zeros
p_diag=diag(p_set)
# compute middle term P-pp' in equation A.1 of Electronic Supplementary Material 2
middle_term<-p_diag-p_set%o%p_set
# pre- and postmultiply with the Xs from the design matrix for the alternatives in this choice set
full_term<-t(design[alternatives,])%*%middle_term%*%design[alternatives,]
# Add contribution of this choice set to the information matrix
info_mat<-info_mat+full_term
} # end of loop over choice sets
#get the inverse of the information matrix (i.e., gets the variance-covariance matrix)
sigma_beta<-solve(info_mat,diag(ncoefficients))
```

Step 7   Sample Size Calculation
          The final step is to calculate the required sample size for the MNL coefficients in our DCE. Hereto we use Eq. 4 (Box 7)

probability that we will find a significant effect given a specific effect size, we also computed the required sample size for the statistical power level $1-\beta$ equal to 0.6, 0.7, and 0.9. Additionally, we also computed the required

---

Box 7

```
# Use the parameter values as effect size. Other values can be used here.
effectsize<-parameters

# formula for sample size calculation is n>[(z_(beta)+z_(1-alpha))*sqrt(Σγκ)/delta]^2
N<-((z_one_minus_beta + z_one_minus_alpha)*sqrt(diag(sigma_beta))/abs(effectsize))^2
# Display results
"required sample size for each coefficient"
N
```

Table 3 DCE design

Choice task	Alternative	Constant	I. Route of drug administration			II. Effectiveness	III. Nausea	IV. Duration	V. Costs
		A	B1	B2	B3	C	D	E	F
1	1	1	1	0	0	5	0	10	120
1	2	1	0	1	0	10	1	1	240
1	3	0	0	0	0	0	0	0	0
2	1	1	0	0	1	5	1	5	720
2	2	1	0	0	0	10	0	10	0
2	3	0	0	0	0	0	0	0	0
3	1	1	0	0	0	25	1	10	240
3	2	1	1	0	0	50	0	1	720
3	3	0	0	0	0	0	0	0	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
16	1	1	0	1	0	10	0	10	720
16	2	1	0	0	1	25	1	1	0
16	3	0	0	0	0	0	0	0	0

alternative 1 = drug treatment alternative I; alternative 2 = drug treatment alternative II; alternative 3 = opt-out alternative; values 0 and 1 in column A mean 'opt-out alternative' and 'drug treatment alternative', respectively; value 1 in columns B1, B2, B3 means 'tablet every week', 'infusion every 4 months', and 'infusion every month', respectively; column C presents how effective (risk reduction of a hip fracture in %) a drug treatment alternative is; values 0 and 1 in column D mean 'no nausea as a side effect' and 'nausea as a side effect', respectively; column E presents the total treatment duration in years; and the values in column F present the out-of-pocket costs (€)

Table 4 Minimum sample size required to obtain the desired power level $1-\beta$ for finding an effect when testing at a specific confidence level $1-\alpha$

$\alpha =$	$1-\beta =$	Constant	I. Route of drug administration			II. Effectiveness	III. Nausea	IV. Duration	V. Costs
		A	B1	B2	B3	C	D	E	F
0.1	0.6	2	28	72	13	2	1	17	3
0.05	0.6	3	43	111	19	2	1	27	4
0.025	0.6	4	58	151	26	3	2	36	6
0.01	0.6	6	79	205	35	5	3	49	8
0.1	0.7	3	39	100	17	2	1	24	4
0.05	0.7	4	56	145	25	3	2	35	6
0.025	0.7	6	73	190	33	4	3	46	7
0.01	0.7	7	96	250	43	6	3	60	10
0.1	0.8	4	53	139	24	3	2	33	5
0.05	0.8	6	73	190	33	4	3	46	7
0.025	0.8	7	93	241	42	5	3	58	9
0.01	0.8	9	119	308	53	7	4	74	12
0.1	0.9	6	78	202	35	5	3	49	8
0.05	0.9	8	102	263	45	6	4	64	10
0.025	0.9	10	125	323	56	7	4	78	13
0.01	0.9	12	154	400	69	9	5	97	16

sample sizes assuming a significance level α of 0.1, 0.025, and 0.01

As can be seen from Table 4, one needs a minimum sample size of 190 respondents with a statistical power of 0.8 and assuming an $\alpha = 0.05$, whether 'injection every 4 months' is significantly different from 'tablet once a month (reference attribute level)' (Table 4,

Table 5 Parameter estimates and precision from an actual discrete-choice experiment study [12] relative to those predicted by the sample size calculations

Attribute	MNL results actual study ($N = 117$)[a]			Predicted results based on 117 subjects	
	Parameter value	SE	95 % CI	SE	95 % CI
Constant (drug treatment)	1.23	0.218	0.81–1.66	0.109	1.02–1.45
Drug administration (base level tablet once a month):					
Tablet once a week	−0.31	0.070	−0.45 to −0.17	0.099	−0.50 to −0.12
Injection every 4 months	−0.21	0.097	−0.41 to −0.02	0.108	−0.43 to −0.01
Injection once a month	−0.44	0.100	−0.64 to −0.25	0.094	−0.63 to −0.26
Effectiveness (1 % risk reduction)	0.03	0.003	0.02–0.03	0.002	0.02–0.03
Side effect nausea	−1.10	0.104	−1.30 to −0.89	0.065	−1.22 to −0.97
Treatment duration (1 year)	−0.04	0.010	−0.06 to −0.02	0.010	−0.06 to −0.02
Cost (€1)	−0.0015	0.0002	−0.002 to −0.001	0.0002	−0.002 to −0.001

CI confidence interval, *SE* standard error

[a] Number of observations 5589 (117 respondents × 16 choices × 3 options per choice, minus 27 missing values), Pseudo $R^2 = 0.185$, log pseudolikelihood $= -1668.7$

column B2). If a smaller sample size of, for example, 111 respondents were to be used and no significant result to be found for this parameter, one has a statistical power of 0.6, assuming an $\alpha = 0.05$, to conclude that respondents do not prefer 'tablet every month' over 'injection every 4 months'. As a proof of principle, we compared the standard errors and confidence intervals from the actual study [12] against the predicted standard errors and confidence intervals. The results showed that they were quite similar (Table 5), which gives further evidence that our sample size calculation makes sense.

5 Discussion

In this paper, we have summarized how researchers have dealt with sample size calculations for health care-related DCE studies. We found that more than 70 % of the health care-related DCE studies published in 2012 did not (clearly) report whether and what kind of sample size method was used. Just 6 % of the health care-related DCE studies published in 2012 used a parametric approach for sample size estimation. Nevertheless, the parametric approaches used were not suitable as a power calculation for determining the minimum required sample size for hypothesis testing for coefficients based on DCEs. To fill in this gap, we explained the analysis needed to determine the required sample size in DCEs from a hypothesis testing perspective. That is, we clarified that the following five elements are needed before such a minimum sample size

can be determined: significance level (α), statistical power level ($1-\beta$), statistical model used in the DCE analysis, initial belief about the parameter values, and the DCE design. An important feature of the resulting sample size formula is that the required sample size tends to grow exponentially. For example, when one wants a certain power level to detect an effect that is 50 % smaller, the required sample will be four times larger.

To build a bridge between theory and practice, we created a generic R-code as a practical tool for researchers to be able to determine the minimum required sample size for coefficients in DCEs. We then illustrate step-by-step how the sample size requirement can be obtained using our R-code. Although the R-code presented in this paper is for MNL only, the theory is also suitable for other choice models, such as the nested logit, mixed logit, scaled-MNL, or generalized-MNL.

Our approach for determining the minimum required sample size for coefficients in DCEs can also be extended to functions of parameters. For example, one might want to know whether patients are willing to pay a specific amount to increase effectiveness by 10 %. In order to test such a hypothesis, confidence intervals for a willingness-to-pay measure are needed. Once how these will be inferred from the limiting distribution of the parameters [42] is determined, Σ_{WTP} (instead of Σ_γ) is known and the required sample size can be computed.

From a practical point of view, in health care-related DCEs, the number of patients and physicians that can be approached is often given, and sometimes rather small. Especially in these cases, our tool could indicate that power will be low. Using efficient designs (striving for small values for $\sum_{\gamma k}$), more alternatives per choice set, or clear

wording and layout are ways to increase the power that is achieved.

The approach presented in this paper can also be used to reverse engineer the power that a specific design has for a given sample size. This can help researchers who find an insignificant result to ensure that they had sufficient power to detect a reasonably sized effect.

6 Conclusion

The use of sample size calculations for healthcare-related DCE studies is largely lacking. We have shown how sample size calculations can be conducted for DCEs when researchers are interested in testing whether a particular attribute (level) affects the choices that patients or physicians make. Such sample size calculations should be executed far more often than is currently the case in healthcare, as under-powered studies may lead to false insights and incorrect decisions for policy makers.

References

1. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.
2. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883–902.
3. Lancaster KJ. A new approach to consumer theory. J Polit Econ. 1966;74(2):132–57.
4. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers in econometrics. New York: Academic Press; 1974. p. 105–42.
5. Reed Johnson F, 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.
6. Louviere J, Hensher DA, Swait JD. Stated choice methods: analysis and application. Cambridge: Cambridge University Press; 2000.
7. Hensher DA, Rose JM, Greene WH. Applied choice analysis: a primer. Cambridge: Cambridge University Press; 2005.
8. Rose JM, Bliemer MCJ. Constructing efficient stated choice experimental designs. Transp Rev. 2009;29(5):587–617.
9. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661–77.
10. Ryan M, Gerards K, Amaya-Amaya M, editors. Using discrete choice experiments to value health and health care. Dordrecht: Springer; 2008.
11. Oteng B, Marra F, Lynd LD, Ogilvie G, Patrick D, Marra CA. Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme. Sex Transm Infect. 2011;87(1):52–7.
12. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HA, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos Int. 2008;19(7):1029–37.
13. Guimaraes C, Marra CA, Gill S, Simpson S, Meneilly G, Queiroz RH, et al. A discrete choice experiment evaluation of patients' preferences for different risk, benefit, and delivery attributes of insulin therapy for diabetes management. Patient Prefer Adherence. 2010;4:433–40.
14. Hiligsmann M, Dellaert BG, Dirksen CD, van der Weijden T, Goemaere S, Reginster JY, et al. Patients' preferences for osteoporosis drug treatment: a discrete-choice experiment. Arthritis Res Ther. 2014;16(1):R36.
15. van Dam L, Hol L, de Bekker-Grob EW, Steyerberg EW, Kuipers EJ, Habbema JD, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. Eur J Cancer. 2010;46(1):150–9.
16. Bessen T, Chen G, Street J, Eliott J, Karnon J, Keefe D, et al. What sort of follow-up services would Australian breast cancer survivors prefer if we could no longer offer long-term specialist-based care? A discrete choice experiment. Br J Cancer. 2014;110(4):859–67.
17. Kimman ML, Dellaert BG, Boersma LJ, Lambin P, Dirksen CD. Follow-up after treatment for breast cancer: one strategy fits all? An investigation of patient preferences using a discrete choice experiment. Acta Oncol. 2010;49(3):328–37.
18. Dixon S, Nancarrow SA, Enderby PM, Moran AM, Parker SG. Assessing patient preferences for the delivery of different community-based models of care using a discrete choice experiment. Health Expect. 2013. doi:10.1111/hex.12096.
19. de Bekker-Grob EW, Hofman R, Donkers B, van Ballegooijen M, Helmerhorst TJ, Raat H, et al. Girls' preferences for HPV vaccination: a discrete choice experiment. Vaccine. 2010;28(41):6692–7.
20. Yeo ST, Edwards RT, Fargher EA, Luzio SD, Thomas RL, Owens DR. Preferences of people with diabetes for diabetic retinopathy screening: a discrete choice experiment. Diabet Med. 2012;29(7):869–77.
21. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. Transportation. 2013;40:1021–41.
22. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2003;2(1):55–64.
23. Huicho L, Miranda JJ, Diez-Canseco F, Lema C, Lescano AG, Lagarde M, et al. Job preferences of nurses and midwives for

taking up a rural job in Peru: a discrete choice experiment. PLoS One. 2012;7(12):e50315.

24. Blaauw D, Erasmus E, Pagaiya N, Tangcharoensathein V, Mullei K, Mudhune S, et al. Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment. Bull World Health Organ. 2010;88(5):350–6.

25. Scott A. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. J Health Econ. 2001;20(3):329–47.

26. Bridges JF, Dong L, Gallego G, Blauvelt BM, Joy SM, Pawlik TM. Prioritizing strategies for comprehensive liver cancer control in Asia: a conjoint analysis. BMC Health Serv Res. 2012;12:376.

27. Bridges JF, Gallego G, Kudo M, Okita K, Han KH, Ye SL, et al. Identifying and prioritizing strategies for comprehensive liver cancer control in Asia. BMC Health Serv Res. 2011;11:298.

28. Bridges JF, Mohamed AF, Finnern HW, Woehl A, Hauber AB. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. Lung Cancer. 2012;77(1):224–31.

29. Bridges JF, Searle SC, Selck FW, Martinson NA. Designing family-centered male circumcision services: a conjoint analysis approach. Patient. 2012;5(2):101–11.

30. Gerard K, Tinelli M, Latter S, Blenkinsopp A, Smith A. Valuing the extended role of prescribing pharmacist in general practice: results from a discrete choice experiment. Value Health. 2012;15(5):699–707.

31. Landfeldt E, Jablonowska B, Norlander E, Persdotter-Eberg K, Thurin-Kjellberg A, Wramsby M, et al. Patient preferences for characteristics differentiating ovarian stimulation treatments. Hum Reprod. 2012;27(3):760–9.

32. Manjunath R, Yang JC, Ettinger AB. Patients' preferences for treatment outcomes of add-on antiepileptic drugs: a conjoint analysis. Epilepsy Behav. 2012;24(4):474–9.

33. Philips H, Mahr D, Remmen R, Weverbergh M, De Graeve D, Van Royen P. Predicting the place of out-of-hours care–a market simulation based on discrete choice analysis. Health Policy. 2012;106(3):284–90.

34. Robyn PJ, Barnighausen T, Souares A, Savadogo G, Bicaba B, Sie A, et al. Health worker preferences for community-based health insurance payment mechanisms: a discrete choice experiment. BMC Health Serv Res. 2012;12:159.

35. Rockers PC, Jaskiewicz W, Wurts L, Kruk ME, Mgomella GS, Ntalazi F, et al. Preferences for working in rural clinics among trainee health professionals in Uganda: a discrete choice experiment. BMC Health Serv Res. 2012;12:212.

36. Tinelli M, Ozolins M, Bath-Hextall F, Williams HC. What determines patient preferences for treating low risk basal cell carcinoma when comparing surgery vs imiquimod? A discrete choice experiment survey from the SINS trial. BMC Dermatol. 2012;12:19.

37. Orme B. Sample size issues for conjoint analysis studies. Sequim: Sawtooth Software Technical Paper; 1998.

38. Johnson R, Orme B. Getting the most from CBC. Sequim: Sawtooth Software Research Paper Series, Sawtooth Software; 2003.

39. Pearmain D, Swanson J, Kroes E, Bradley M. Stated preference techniques: a guide to practice. 2nd ed. Steer Davies Gleave and Hague Consulting Group. 1991.

40. Pedersen LB, Kjaer T, Kragstrup J, Gyrd-Hansen D. Do general practitioners know patients' preferences? An empirical study on the agency relationship at an aggregate level using a discrete choice experiment. Value Health. 2012;15(3):514–23.

41. Bliemer MCJ, Rose JM. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transp Res B Methodol. 2010;44(6):720–34.

42. de Bekker-Grob EW, Rose JM, Bliemer MC. A closer look at decision and analyst error by including nonlinearities in discrete choice models: implications on willingness-to-pay estimates derived from discrete choice data in healthcare. Pharmacoeconomics. 2013;31(12):1169–83.