

## Title

Understanding uptake of digital health products: Methodology tutorial for a Discrete Choice Experiment using a Bayesian efficient design

## Authors

Dorothy Szinay<sup>1</sup>, Rory Cameron<sup>2</sup>, Felix Naughton<sup>1</sup>, Jennifer A. Whitty<sup>2</sup>, Jamie Brown<sup>3,4</sup>, Andy Jones<sup>2</sup>

<sup>1</sup>Behavioural and Implementation Science Group, School of Health Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7UL, UK

<sup>2</sup>Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

<sup>3</sup>Department of Behavioural Science and Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK

<sup>4</sup>SPECTRUM Consortium, London, UK

Corresponding author:

Dorothy Szinay,

Behavioural and Implementation Science Group, School of Health Sciences, University of East Anglia

Norwich Research Park, Norwich NR4 7UL, UK

Email: [d.szinay@uea.ac.uk](mailto:d.szinay@uea.ac.uk)

Proposed journal: JMIR

Word count: 4,936

## Abstract

Understanding the preferences of potential users of digital health products is beneficial for digital health policy and planning. Stated preference methods could help elicit individuals' preferences in the absence of observational data. A discrete choice experiment (DCE) is a commonly used stated preference method; a quantitative methodology that argues that individuals make trade-offs when engaging in a decision by choosing an alternative of a product or service that offers the greatest utility, or benefit. This methodology is widely used in health economics in situations where revealed preferences are difficult to collect but is much less used in the field of digital health. This article outlines the stages involved in developing a discrete choice experiment. As a case study, it uses the application of a DCE for revealing preferences in targeting the uptake of smoking cessation apps. It describes the establishment of attributes, the construction of choice tasks of two or more alternatives, and the development of the experimental design. This tutorial offers a guide for researchers with no prior knowledge of this research technique.

Keywords: Discrete Choice Experiment, stated preference methods, mhealth, digital health, quantitative methodology, uptake, engagement

## Introduction

Understanding how the public value different aspects of digital health tools, such as smoking cessation or physical activity apps, can help providers of the tools to identify functionality that is important to users, which may improve uptake (i.e., selection, download and installation of apps) [1]. This is important because uptake of digital tools is generally low. More information regarding the preferences of users when selecting a digital health tool, for example via an app store, may allow providers to present their products in such a way that may increase their uptake. However, pragmatic challenges, such as examining how each potentially modifiable aspect of a digital health product (e.g. presentation, design and features that it offers) or intervention design will impact preference or choice of uptake, often mean this is not feasible or practical [2]. Therefore, increasing attention is being paid towards stated preference methods to understand preferences when designing digital health products and services, with examples including COVID tracing apps [3,4], sun protection apps to prevent skin cancer [5], and the uptake of health apps in general [6].

Stated preference methods are survey-based methods aiming to elicit individuals' preferences on a specific behaviour, particularly those that are not well understood. The most widely used type of stated preference method is the discrete choice experiment (DCE) [7]. Louviere and Hensher (1982) and Louviere and Woodworth (1983) originally developed DCEs to study the marketing and economics of transport, and the fields of psychology and economics have profoundly influenced the DCE methodology since it was developed [8]. In recent years, DCEs have been increasingly employed in health and health care settings [9,10], as well as in addiction research [11] and digital health [4-6]. The increasing number of DCEs in digital health highlights their potential although they are currently underutilised.

Discrete choice differentiates from other stated preference methods in the way that responses are elicited [12]. The DCE uses a survey-based experimental design where participants are presented with a series of hypothetical scenarios. In these scenarios, participants are shown situations, known as *choice tasks*. Attempting to mimic real-world decision-making, in each choice task participants then have to choose a product or a service from two or more options, known as *alternatives* [13]. Each alternative consists of a set of characteristics, known as *attributes*, with at least two types, known as *attribute levels* [13]. Participants are asked to choose a preferred alternative in each choice task, which allows researchers to quantify the relative strength of preferences for improvements in certain attributes [8,14].

The outputs from statistical models developed using DCE data can be beneficial for estimating uptake of new products or services, including digital health tools, where

observational data is not available or is difficult to obtain otherwise [15,16]. Lack of observational data often implies a requirement to seek scientific views and comments from experts, to generate predictions of a target behaviour [17]. However, DCEs can provide an empirical alternative to expert opinions while accounting for possible interactions between attributes (e.g. design of a product and brand name), which are otherwise often ignored [18].

In our research we wanted to understand how to present health apps on curated health app portals to increase their uptake. This paper describes the development of a DCE in digital health that aimed to elicit potential user preferences on smoking cessation app uptake. It explains how the attributes and their levels are selected and describes the construction of choice tasks and the experimental design. The study protocol of the research this paper is based on is registered on the Open Science Framework [19].

### The development of a discrete choice experiment

The development of the DCE should follow published recommendations, including the checklist for good research practices [9], guides on the development of a DCE [13,20], recommendations on how to construct the experimental design [20-24], and which statistical methods can be used [25].

### Establishing attributes

An important step in designing a DCE is the identification of the relevant attributes for the subject matter. Attributes in a DCE can be quantitative, such as cost, or qualitative, such as the design of a product [26]. The identification of attributes is typically based on primary and secondary data collection to ensure that the DCE is tailored to the study setting [13]. It should ideally commence with a literature review which will inform qualitative research to identify relevant attributes [27]. Although there is no set limit on the number of attributes that can be included in a DCE, to ensure that the cognitive load of the participants is manageable, it should be less than ten [13] with a general expectation to include five to seven attributes [28].

Our DCE was based on a comprehensive systematic review investigating factors influencing the uptake and engagement with health and wellbeing smartphone apps [29] and a qualitative research component that consisted of a think-aloud and interview study to examine further the previously identified factors or attributes [30]. The importance of qualitative research lies in ensuring inclusion of attributes that are relevant to most participants [26]. Of the 14 factors initially identified as being relevant for the uptake of health and wellbeing apps, 5 were retained and included in the DCE: *the monthly price of the app*, *who developed the app*, *the star ratings of the app*, *the description of the app* and *images shown*. These factors were chosen due to their perceived importance during our previous

qualitative research and for pragmatic reasons including how easily measurable and presentable they were within a DCE.

An important step in designing a DCE is in ensuring the content validity of the instrument: the identification of the relevant attributes for the subject matter. These factors were chosen due to their perceived importance during our previous qualitative research and for pragmatic reasons including how easily measurable and presentable they were within the DCE. Following administration of the survey, methods are available for the measurement and assessment of the content validity of the instrument, although their use is not widely reported [31].

#### *Establishing attribute levels*

The next step is to establish the attribute levels. The level of an attribute must also be of a range that ensures a trade-off between attributes. A trade-off is defined as an exchange in which a participant gives up some amount of one attribute to gain more of another. It has been suggested that increasing the number of levels for an attribute increases the relative importance of that attribute [32], and that imbalance of numbers of levels across attributes raises the importance of the attributes with higher levels [33]. Yang and colleagues have suggested a balance exists between simpler designs with lower numbers of levels, which reduce respondent burden (and consequently measurement error) and are useful for identifying attribute rankings; and more complex designs with higher levels (and higher statistical precision) and are more sensitive to identifying tradeoffs between attributes [33]. Based on this, and the commonly adopted practices in the research field, we aimed to include at least three levels for each attribute.

If a range is not suitable, participants might consider the differences between levels unimportant [26]. For example, the difference of the star ratings of 4.8 and 4.7 of a smoking cessation app are not as relevant as a difference of 4.8 and 4. In our research, to refine the attribute levels, a survey was conducted with 34 participants. In the survey, the levels of two attributes we unsure of, the monthly price of the app and the ratings, were carefully considered so that the levels of these two attributes were specified at a sufficiently wide range that the difference between the levels would likely make a difference in response. When a range is not wide enough, there is a risk that participants could ignore the attributes because they judge the difference between levels to be insignificant [20]. See Figure 1 for the final list of attributes and levels included in our DCE.

Attributes	Attribute levels
1. The monthly price of the app	<ul style="list-style-type: none"> <li>£0</li> <li>£2.99</li> <li>£5.99</li> <li>£8.99</li> </ul>
2. Who developed the app	<ul style="list-style-type: none"> <li>Doesn't say</li> <li>'Mhealth Essentials Ltd.'</li> <li>'NHS Digital'</li> </ul>
3. The ratings of the app	<ul style="list-style-type: none"> <li>Doesn't show</li> <li>3.2 ★★★★☆</li> <li>4.0 ★★★★☆</li> <li>4.8 ★★★★★</li> </ul>
4. App description	<ul style="list-style-type: none"> <li><b>Generic</b>, to create a rough idea of what the app is about without getting into details of app features</li> <li><b>Short</b> with some details about app features</li> <li><b>Long</b> and detailed description of the app and its features</li> </ul>
5. Images	<ul style="list-style-type: none"> <li>Shows the logo of the app</li> <li>Shows the screenshot(s) of the app</li> <li>Shows the logo and screenshot(s) of the app</li> </ul>

Figure 1. Attributes and attribute levels in our DCE.

#### Choice tasks

Once the attributes and their levels are identified, the decision to develop 'full-profile' or 'partial-profile' tasks with or without an opt-out option needs to be made. Full-profile refers to the display of all five attributes in both alternatives in each choice-set. A partial-profile DCE will not present certain attributes for certain alternatives. For example, if a DCE was used to investigate the trade-off between a higher number of attributes (e.g. a total of nine attributes), it could be beneficial to limit the number of attributes shown at one time (e.g. five attributes) to limit participant cognitive load. Five attributes is generally considered low enough to complete a full-profile choice task which consequently maximises information about trade-offs [34]. Hence, in our research, we applied a full profile DCE.

A neutral option ('Neither of these two'), known as an opt-out alternative, was included in addition to selecting alternative apps. The opt-out option has the potential to make the choices more realistic [35] by simulating a real-world context where individuals can exercise their right not to take up an app, given the apps on offer [20]. In our DCE, a participant had the option to choose or reject the hypothetical uptake of a smoking cessation app. However, where a participant selects the opt-out option, no information is provided on how they trade-off attribute levels or alternatives [13]. In some situations, a *forced-choice* scenario can be

included, where participants who chose an opt-out option are prompted to make a choice regardless. An example of a scenario with an opt-out option is shown in Figure 2.

You wish to quit smoking, and you decide to select a smartphone app to do that. Please look at the options carefully, and decide on which app (App 1 or App 2) you think you would likely want to download and use to help you quit smoking.

You could also choose 'Neither of these two' if you do not like either option and would not choose to download either app.

Take your time to make a decision. Please, select an option and click on the arrow to continue.

	<b>App 1</b>	<b>App 2</b>
<b>The monthly price of the app</b>	£8.99	£0
<b>Who developed the app</b>	Mhealth Essentials Ltd.	NHS Digital
<b>The ratings of the app</b>	4.8 ★★★★★	4.0 ★★★★☆
<b>App description</b>	Generic, to create a rough idea of what the app is about without getting into details of app features	Short with some details about app features
<b>Images shown</b>	Logo and screenshot(s) of the app	Logo of the app

App 1

App 2

Neither of these two

Figure 2. An example of a scenario with an opt-out option used in our discrete choice experiment.

### Experimental design

An experimental design is a systematic method of generating the choice sets that are presented to respondents. This one enables the specification of the choice sets that respondents see, with the objective of obtaining a high quality data set [7]. When creating the experimental design, there are several aspects that need to be taken into consideration including: 1) the analytical model specification; 2) whether the aim is to estimate main effects only or interaction effects as well; 3) whether the design is labelled or unlabelled, 4) the number of choice tasks and blocking options to be used, 5) which type of design of the choice matrix to use (e.g. full factorial or fractional factorial; orthogonal or efficient design), and 6) how the attribute level balance is achieved. These are now considered.

### Analytical model specification

The first step in the generation of an experimental design is to specify the analytical model to estimate the parameters of the DCE. This step is an important component of choosing the type of choice matrix design, described later in this paper. The approach selected here needs to be accounted for when generating the structure of the experimental design.

A discrete choice model describes the probability that an individual will choose a specific alternative. This probability is expressed as a function of measured attribute levels specific to the alternative and of characteristics of the individual making the choice. This probability is represented by the dependent variable (the *choice variable*), which indicates the choice made by participants [8]. In this modelling framework the attributes are the independent variables [8,13].

As part of the analytical model specification, knowing what type of statistical analysis will be used is key. Data analysis involves regression modelling in a random utility framework [8]. The random utility model conventionally used is also based on Lancaster's theory of consumer demand [36] which together assume that individuals make trade-offs when making a decision, and would choose an option that offers the greatest utility [37], determined by how much importance the individual places on the attributes associated with the product [38].

The Multinomial Logit (MNL) model has been previously described as the 'workhorse' of DCE estimation [39,40], and it typically serves as a starting point for basic model estimation (although alternative models, such as probit, may be used). It is important to note that MNL requires some important assumptions and limitations; for example independence of irrelevant alternatives, homogeneity of preferences, and independence of observed choices [41,42]. Extensions of MNL (e.g. nested logit, mixed logit, and latent class models) may be employed to account for these limitations [40,41].

Based on the model specified in our DCE, the underlying utility function for alternative  $j$  [39] is shown in Box 1:

$$U_j = (\beta_{cost} * X_{j_{cost}}) + (\beta_{developer} * X_{j_{developer}}) + (\beta_{ratings} * X_{j_{ratings}}) + (\beta_{description} * X_{j_{description}}) + (\beta_{images} * X_{nj_{images}}) + \epsilon$$

Note:

- 1)  $U$  the overall utility derived from alternative  $j$
- 2)  $\beta$  is the coefficient attached to  $X_j$  estimated in analysis and represent the part-worth utility attached to each attribute level
- 3)  $\epsilon$  is the random error of the model, in other words, the unmeasured factors influencing variation of preferences

Box 1. The utility function used in our DCE research

### *Main effects or interaction effects*

The next step in model specification is deciding whether main effects or interaction effects will be investigated. Main effects, the most commonly used, investigate the effect of each attribute level on the choice variable. The effect on the choice variable gained by combining two or more attribute levels (e.g. app developer and the app's monthly cost) refers to an interaction effect [13]. In our DCE given the novel nature of the research in the uptake of health apps and the lack of empirical evidence to suggest the presence of potential interactions between attributes, we decided to only look at main effects.

### *Labelled or unlabelled*

In a labelled experiment, the alternatives are specific and different (e.g. smartphone app-based smoking cessation intervention vs website-based smoking cessation intervention) and alternative specific attributes could be used (e.g. some attributes relevant only for apps and others for websites). This is in contrast to an unlabelled experimental design, where the alternatives are unspecified (e.g. smoking cessation app alternative 1 vs smoking cessation app alternative 2) and also must have the same attributes. Given a DCE model estimates parameters for each of the alternatives being considered, these alternative specific parameters must be included in the structure of the experimental design (described in the next section) in a labelled experiment; in an unlabelled experiment, because they are arbitrary, they are excluded [23,43,44]. In health economics, the unlabelled approach is the most common. In our DCE the unlabelled approach was deemed to be logical here as we were comparing different presentations of the same app. Therefore, our DCE design applied an unlabelled approach.

### *Generation of the structure of the experimental design*

Once the model is specified, the structure of the experimental design can be generated. For this stage hypothetical alternatives are generated and combined to form choice tasks, based on the chosen attributes and their levels. Several different software packages may be used to generate the experimental design of a DCE, such as Ngene, SAS, SPEED, SPSS, Sawtooth. For our DCE, Ngene software was used [45].

### *The number of choice tasks and blocking*

The next step in the generation of an experimental design is to decide on the choice task and blocking. In order to minimise respondent and cognitive burden, and the risk of participants losing interest during the DCE task, consideration must be paid to the target population, the number of tasks, and their complexity [13]. The higher the number of attributes, alternatives and choice tasks, the higher the task complexity [20]. The literature suggests that a feasible limit is 18 choice sets per participant [46,47]. In the review by

Marshall and colleagues, most studies included between 7 and 16 choice sets [28]. In our DCE we administered 12 choice tasks per participant, which were deemed to be a number low enough to avoid excessive cognitive load but high enough to establish sufficient statistical precision.

We developed forty-eight choice tasks and blocked into four survey versions (12 choice tasks for each). Each block represented a separate survey and participants were randomly assigned to one of the four survey versions. Blocking is a technique widely used in DCEs to reduce cognitive burden, by partitioning large experimental designs into subsets of equal size, and thereby reducing the number of choice tasks that any one respondent is required to complete [48]. Blocks were generated in Ngene software, which allows for the minimisation of the average correlation between the versions and attributes' levels [49]. For the blocking to be successful, the number of choice tasks included in one block must be divisible by the number of the attribute levels; in our DCE, attributes had either three or four levels.

It is noteworthy, that in order to undertake the sample size calculation, it is crucial to know the number of alternatives per choice set, the largest number of levels of any attribute (for DCEs looking at main effects only) or the largest level of any two attributes (for DCS looking at interaction effects) and the number of blocks [39]. Therefore, DCEs using blocking require a larger sample size [48].

#### Type of choice matrix design

Depending on the number of attributes and their levels, a full-factorial or fractional factorial design can be applied. A full factorial design would include all possible combinations of the attributes' levels and allow the estimation of all main effects and interaction effects independently of one another [20]. However, this type of design is often considered impractical due to the high number of choice tasks required [20]. To illustrate this, the formula of calculation of the possible unique choice alternatives for a full factorial design is:  $L^A$ , where L represent the number of levels and A the number of attributes [40]. If the attributes in the DCE have a different number of levels these need to be calculated separately and multiplied together. In order to reduce response burden, in our DCE we generated a fractional factorial design in Ngene [45], representing a sample of possible alternatives from the full factorial design. In this way we were able to reduce the total 432 alternatives in the full design (given by  $L^A = 4^2 \times 3^3$ ), to a fractional sample of 96 alternatives, arranged in 48 choice pairs.

Systematic approaches for generation of fractional factorial designs may further subset into orthogonal design and efficient design. An orthogonal design is a column-based design

based on orthogonal arrays which present properties of orthogonality (attributes are statistically independent of one another) and level balance (levels of attributes appear an equal number of times), and does not introduce correlation between the attributes [39]. An orthogonal array is an optimal design that is often used for DCEs examining main effects when the number of attributes and their levels are small.

For studies with five or more attributes with two or more levels, an orthogonal design may not be practical. There has therefore been a recent change in thinking toward a nonorthogonal and statistically more efficient design [39]. When perfect orthogonality and balance cannot be achieved or are not desirable, an efficient design can be applied [20]. In contrast to an orthogonal design, an efficient design aims to increase precision of parameter estimates for a given sample size (i.e. minimising the standard errors of the estimated coefficients), while allowing some limited correlation between attributes. The most widely used efficiency measure is *D-error* which may be easily estimated using various software packages such as Ngene, and refers to the efficiency of the experimental design in extracting information from respondents [22]. Experimental designs generated using this approach are known as D-efficient designs. A D-efficient experimental design is also recommended to maximise statistical efficiency and minimise the variability of parameter estimates [7].

An efficient design requires that known prior information about the parameters (known as 'priors') are made available to the algorithm, and also requires the analyst to specify the analytical model specification, as described previously. Depending on what information is available, one of three types of D-efficient design can be generated [22].

- 1) *D<sub>z</sub>-efficient* design (*z* stands for zero priors) - if no prior information about the magnitude or directions of the parameters are available (*D<sub>z</sub>-efficient* design is an orthogonal design). This design assumes the parameters are zero.
- 2) *D<sub>p</sub>-efficient* design (*p* stands for priors) - assumes a fixed, certain value and direction for the parameters
- 3) *D<sub>b</sub>-efficient* design (*b* stands for Bayesian) – A Bayesian approach whereby the parameter is not known with certainty, but may be described by its probability distribution

Best practice is to pilot the DCE. For the pilot phase, there is limited information available and using *D<sub>z</sub>-efficient* or *D<sub>p</sub>-efficient* design is sensible. In our DCE we chose to apply a *D<sub>p</sub>-efficient* design as the direction of priors of the app was known from the previously conducted survey narrow down the attribute levels and to provide prior estimates of the parameters for the attribute levels. For example, we knew that a trusted organisation will

likely positively influence uptake and cost estimated negatively so. The direction of priors was assumed to be a very small near zero negative or positive value for the design.

The pilot phase provided estimation that we used to generate a *D<sub>b</sub>-efficient* design for the final DCE. It is noteworthy that when the parameter priors are different from zero, the efficient design generated produces smaller prediction errors than the orthogonal designs [22,50,51]. Hence, a D-efficient design will outperform an orthogonal design, and, (given reliable priors) a *D<sub>p</sub>-efficient* design will outperform a *D<sub>z</sub>-efficient* design [22]. Further, when reasonable assumptions about the distributions are made, a *D<sub>b</sub>-efficient* design will outperform a *D<sub>p</sub>-efficient* design. Therefore, it may be advisable to start piloting with a *D<sub>p</sub>-efficient* design and to generate a *D<sub>b</sub>-efficient* design for the final DCE. The DCE literature provides a detailed and more comprehensive description of the orthogonal and efficient designs [22], and approximation of Bayesian efficient design [24].

#### Attribute level balance in the model

The attribute level balance aims to ensure all attribute levels ideally appear an equal number of times in the experimental design. The allocation of the attribute levels within the experimental design can affect statistical power; if a certain level is under-represented in the choice sets generated, then the coefficient for that level cannot be easily estimated. How attributes levels are distributed is therefore an important consideration when designing the choice sets. Dominant alternatives, where all attribute levels of one alternative are more desirable than all attribute levels in the other, do not provide information of how trade-offs are made, as individuals usually would select the dominant alternative. Therefore, avoiding dominant alternatives in the experimental design is important and can be achieved by consulting the software manual to ensure the correct algorithm is used. The syntax used in Ngene to generate choice sets of the pilot phase and more information about the algorithm used can be accessed on the Open Science Framework [19].

#### Piloting the DCE and generating the Bayesian design

As well as providing estimations for the choice matrix design described above, piloting offers an opportunity to ensure that the information is presented clearly, and that the choices are realistic and meaningful. It also provides insight into how cognitively demanding it is for respondents to complete. This can be achieved by gathering feedback on the survey completion process. The findings of the pilot may suggest that the DCE needs to be amended, such as reducing the number of choice sets or the number of attributes, so that the responses are a better reflection of participants' preferences and improve the precision in the parameter estimates [13].

There is no formal guidance on how large the pilot sample should be, this is largely guided by budget and complexity of the experimental design. Accuracy of the priors will improve with increasing sample, but as few as 30 responses may be sufficient to generate useable data [45]. In our pilot study conducted with 49 individuals, feedback from participants suggested that with the initial order of the attributes there was a tendency to ignore the last two attributes, the app description and images, the most text heavy attributes. This may have compromised the examination of the relative importance of those two attributes (description and images of the app). Therefore, we decided to change the final order of the attributes from 1) *monthly price of the app*, 2) *the ratings of the app*, 3) *who developed the app*, 4) *the description* and 5) *images shown*, to the one listed in Figures 1 and 2. The longest completion time for the survey was under 12 minutes. Thus, we concluded that the number of choice tasks did not need to be reduced.

In our research, the data from the pilot phase was analysed using the freely available Apollo package in R [52]. The coefficients and their standard errors from the output were used as priors to generate the final choice sets using the Bayesian efficient design following the steps described previously. The syntax used in R used to analyse the pilot data and that used to generate the Bayesian efficient design in Ngene can be accessed on the Open Science Framework [19].

#### Internal validity

Assessing the internal validity of a DCE can help with understanding the consistency and trade-off assumptions made by participants [53]. There are several ways to examine the internal validity of a DCE. For example, in the *stability validity test*, a choice task would be repeated later in the sequence to investigate the consistency of the participants' decision, whether the respondent would choose the same alternative [53]. Another way to test internal validity is the *within-set dominated pairs* type of internal validity in which one alternative is a dominant alternative in which all attributes are of the most desirable. The choice sets designed to measure the internal validity are excluded from the analysis. There are several internal validity tests that are built in software packages such as MATLAB [53], although these can be produced manually as well. In our research we used the stability validity test to check the internal validity by repeating a randomly generated choice task (in our case it was the fourth). Therefore, participants were shown 12 choice tasks, plus an additional 'hold-out', task. The data from the randomly generated hold-out task was excluded from the analysis.

While internal validity checks provide some measure of data quality, it should be noted that answering a repeat choice inconsistently is not a violation of random utility theory [54]. Furthermore, there is no consensus on what to do with the data from responses that 'fail'

validity tests. Following the advice of Lancsar and Louviere (2006) we did not exclude participants who 'failed' the internal validity check as that may cause statistical bias or affect statistical efficiency [55]. However, we reported data on internal validity to enable the reader to make a judgement on likely biases.

All additional study materials used in our example, including the full dataset and the results of the DCE, can be accessed on Open Science Framework [19].

## Discussion

This paper describes the development of a DCE, following the stages required to establish attributes and their levels, construct choice-tasks, define the utility model, decide on labelled and unlabelled choices to apply, decide on the number of choice tasks that need to be generated, as well as make decisions on the structure of the experimental design, how to achieve attribute level balance, to assess the internal model validity, and to pilot test. In doing so, the intention is to advance methodological awareness of the application of stated preference methods in the field of digital health, as well as to provide researchers with an overview of their application using a case study of a DCE of smoking cessation app uptake.

Although DCEs are widely used to understand patient and provider choices in healthcare [8,10,15,56], they have only recently started to gain popularity in digital health [4-6], and as such represent an underused approach in digital health. With the growing evidence of the benefit of digital health initiatives, there are clear benefits to widening the application of DCEs so that they may more routinely inform digital health development, digital tool presentation, and most importantly to predict uptake and engagement with digital products. Whilst several attempts have been made to measure engagement with digital tools using a wide range of methodologies [57-59], the insights we have from them that can be translated to uptake are limited. One plausible explanation is that uptake of digital tools is difficult to empirically measure.

DCEs bring several benefits to help overcome the issue of measuring uptake in digital health or in other areas where the measurement of the predictors of uptake in a good or service is required. For example, as illustrated by the case study here, they enable the researcher to gain measurable insights into situations where quantitative measures are hard to otherwise obtain, such as the factors impacting the uptake of health apps on curated health app portals. A DCE also helps to quantify preferences to support more complex decisions [60]. An example would be the consideration of how to plan the development of an app that would provide appealing looks or features that would promote uptake. The DCE methodology is also considered to be a convenient approach to investigate the uptake of new interventions, including digital health interventions [39], for example digital behaviour change interventions

using health and wellbeing smartphone app. Therefore, DCEs can be used in hypothetical circumstances, enabling the measurement of preferences for a potential policy change or digital health system change before it is implemented [13], such as the recent investigation of the uptake of a COVID-19 test and trace health app [3,4]. The experimental nature of the DCE also means that participants' preferences can be recorded based on controlled experimental conditions where attributes are systematically varied by researchers to provide insight into the marginal effect of attribute changes on individuals' choices [61].

Despite their benefits, the application of DCEs present several challenges. As with all expressed preference methodologies, the hypothetical nature of the DCE choice-set raises concerns about external validity and the degree to which real-world decisions might equate to those made by study participants under experimental conditions, a phenomenon known as the intention-behaviour gap [62]. As such, participants may believe they would choose a scenario presented and described in a choice task, but in real-life there might be other factors that would influence their behaviours, such as the aesthetics of the app [29]. This limitation can at least partially be overcome by developing convincing and visually appealing choice tasks. Nevertheless, to date there has been limited progress in testing for external validity due to the difficulty in investigating preferences in the real world [39]. Indeed, a recent systematic review of the literature on DCEs in health care reported that only 2% of the included studies (k=7) reported details of the investigation of external validity [48], whilst an earlier systematic review and meta-analysis (k=6) found DCEs have only a moderate level of accuracy in predicting behaviours of health choices [63]. To our knowledge, no study has been published that investigates the external validity of a DCE developed in digital health. One potential opportunity to undertake some testing would be through a curated health app portal, where the same health app is presented in two or more different ways. With the help of website analytics actual user behaviour could be measured in this situation.

A final significant concern associated with the use of a DCE is that any single choice set is unlikely to be able to present the user with all relevant attributes, regardless of how well it has been developed [63]. Choosing the most relevant attributes to test in a DCE, therefore, requires comprehensive preparatory research, which can lengthen the time required to undertake the development phase of any piece of work.

In summary, DCEs have significant potential in digital health research, and can serve as an important decision-making tool in a field where observational data is lacking. It is hoped that the content of this paper provides a useful introduction and a guide to those interested in developing such experiments in digital health.

## Acknowledgements

We are grateful to two experts in DCEs, Professor Michiel Bliemer, from the University of Sydney, Australia, a co-developer of the Ngene software, and Professor Stephane Hass from the University of Leeds, United Kingdom, a co-developer of the Apollo package in R software, for their advice on the syntax used to generate the choice tasks in Ngene, and on the code used in the Apollo package.

## Authors' Contributions

DS prepared the manuscript. All authors reviewed the draft for important intellectual content and approved the final version.

## Conflicts of Interest

None declared

## References

1. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios C-M, et al. Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. *International Journal of Technology Assessment in Health Care*. 2011;27(4):369-75. PMID: 22004779
2. Hall J, Viney R, Haas M, Louviere J. Using stated preference discrete choice modeling to evaluate health care programs. *Journal of Business Research*. 2004;57(9):1026-32. PMID: 12112494
3. Wiertz C, Banerjee, A., Acar, O.A., Ghosh, A. . Predicted Adoption Rates of Contact Tracing App Configurations - Insights from a Choice-Based Conjoint Study with a Representative Sample of the UK population. SSRN. 2020. PMID:
4. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mólken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. *JMIR Mhealth Uhealth*. 2020;8(10):e20741. PMID: 32795998
5. Nittas V, Mütsch M, Puhan MA. Preferences for Sun Protection With a Self-Monitoring App: Protocol of a Discrete Choice Experiment Study. *JMIR Res Protoc*. 2020;9(2):e16087. PMID: 33245282
6. Leigh S, Ashall-Payne L, Andrews T. Barriers and Facilitators to the Adoption of Mobile Health Among Health Care Professionals From the United Kingdom: Discrete Choice Experiment. *JMIR Mhealth Uhealth*. 2020;8(7):e17704. PMID: 32628118
7. 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. PMID: 23337210
8. Spinks J, Chaboyer W, Bucknall T, Tobiano G, Whitty JA. Patient and nurse preferences for nurse handover—using preferences to inform policy: a discrete choice experiment protocol. *BMJ Open*. 2015;5(11):e008941. PMID: 26560060
9. Bridges JFP, 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 in Health*. 2011;14(4):403-13. PMID: 21669364
10. 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. PMID: 25005924
11. Kotnowski K, Fong GT, Gallopel-Morvan K, Islam T, Hammond D. The Impact of Cigarette Packaging Design Among Young Females in Canada: Findings From a Discrete Choice Experiment. *Nicotine Tob Res*. 2016;18(5):1348-56. PMID: 26014454
12. Lambooi MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YWM, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC medical research methodology*. 2015;15(1):19. PMID: 25887890
13. Mangham LJ, Hanson K, McPake B. How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning*. 2008;24(2):151-8. PMID: 19112071

14. Trapero-Bertran M, Rodríguez-Martín B, López-Bastida J. What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. *PLOS ONE*. 2019;14(7):e0219905. PMID: 31318926
15. Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. *Health Economics*. 2002;11(5):457-65. PMID: 12112494
16. Fiebig DG, Knox S, Viney R, Haas M, Street DJ. Preferences for new and existing contraceptive products. *Health Economics*. 2011;20(S1):35-52. PMID: 21809412
17. Terris-Prestholt F, Hanson K, MacPhail C, Vickerman P, Rees H, Watts C. How Much Demand for New HIV Prevention Technologies Can We Really Expect? Results from a Discrete Choice Experiment in South Africa. *PLOS ONE*. 2014;8(12):e83193. PMID: 24386160
18. Terris-Prestholt F, Quaife M, Vickerman P. Parameterising User Uptake in Economic Evaluations: The role of discrete choice experiments. *Health Economics*. 2016;25(S1):116-23. PMID: 26773825
19. Szinay D, Rory, C., Jones, A., Whitty, J., Chadborn, T., Jamie, B., Naughton, F. Eliciting adult smokers' preferences for the uptake of smoking cessation apps: A Discrete Choice Experiment (2021, March 12). Retrieved from [osf.io/5439x](https://osf.io/5439x). 2021.
20. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *PharmacoEconomics*. 2008;26(8):661-77. PMID: 18620460
21. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher 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 in Health*. 2013;16(1):3-13. PMID: 23337210
22. Rose JM, Bliemer MCJ. Constructing Efficient Stated Choice Experimental Designs. *Transport Reviews*. 2009;29(5):587-617. PMID:
23. de Bekker-Grob EW, Hol L, Donkers B, van Dam L, Habbema JDF, van Leerdam ME, et al. Labeled versus Unlabeled Discrete Choice Experiments in Health Economics: An Application to Colorectal Cancer Screening. *Value in Health*. 2010;13(2):315-23. PMID: 19912597
24. Bliemer MCJ, Rose JM, Hess S. Approximation of bayesian efficiency in experimental choice designs. *Journal of Choice Modelling*. 2008;1(1):98-126. PMID:
25. Hauber AB, González JM, Groothuis-Oudshoorn CGM, 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 in Health*. 2016;19(4):300-15. PMID: 27325321
26. Kløjgaard ME, Bech M, Søgaard R. Designing a Stated Choice Experiment: The Value of a Qualitative Process. *Journal of Choice Modelling*. 2012;5(2):1-18. PMID:
27. Buchanan J, Blair E, Thomson KL, Ormondroyd E, Watkins H, Taylor JC, et al. Do health professionals value genomic testing? A discrete choice experiment in inherited cardiovascular disease. *Eur J Hum Genet*. 2019;27(11):1639-48. PMID:
28. Marshall D, Bridges JF, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008. *Patient*. 2010;3(4):249-56. PMID: 22273432

29. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review. *J Med Internet Res*. 2020. PMID: 32348255
30. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the Uptake of Health and Well-being Apps and Curated App Portals: Think-Aloud and Interview Study. *JMIR Mhealth Uhealth*. 2021;9(4):e27173. PMID: 33904827
31. Rakotonarivo OS, Schaafsma M, Hockley N. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *Journal of Environmental Management*. 2016;183:98-109. PMID: 27576151
32. Ratcliffe J, Longworth L. Investigating the structural reliability of a discrete choice experiment within health technology assessment. *Int J Technol Assess Health Care*. 2002;18(1):139-44. PMID: 12025111
33. Yang JC, Reed SD, Hass S, Skeen MB, Johnson FR. Is Easier Better Than Harder? An Experiment on Choice Experiments for Benefit-Risk Tradeoff Preferences. *Med Decis Making*. 2021;41(2):222-32. PMID: 33463397
34. Mühlbacher A, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. *Appl Health Econ Health Policy*. 2016;14(3):253-66. PMID: 26992386
35. Watson V, Becker F, de Bekker-Grob E. Discrete Choice Experiment Response Rates: A Meta-analysis. *Health Economics*. 2017;26(6):810-7. PMID: 27122445
36. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer Netherlands; 2008. p. 13-46. PMID: 18644444
37. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P., Ed., *Frontiers in Econometrics*. In: Press A, editor. 1973. p. 105-42. PMID: 10902721
38. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data. *Social Science & Medicine*. 2011;72(10):1717-27. PMID: 21530040
39. 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. PMID: 22223558
40. Hensher DA RJ, Green WH. *Applied choice analysis: a primer*. . In: Press CU, editor. 2005. PMID: 16032721
41. Hensher DA, Greene WH. The Mixed Logit model: The state of practice. *Transportation*. 2003;30(2):133-76. PMID: 12602721
42. Train KE. *Logit. Discrete Choice Methods with Simulation*. 2 ed. Cambridge: Cambridge University Press; 2009. p. 34-75. PMID: 19402721
43. Kruijshaar ME, Essink-Bot M-L, Donkers B, Looman CWN, Siersema PD, Steyerberg EW. A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. *BMC medical research methodology*. 2009;9(1):31. PMID: 19402721
44. Jin W, Jiang H, Liu Y, Klampfl E. Do labeled versus unlabeled treatments of alternatives' names influence stated choice outputs? Results from a mode choice study. *PLOS ONE*. 2017;12(8):e0178826. PMID: 28806764
45. ChoiceMetrics (2012) *Ngene 1.1.1 User Manual & Reference Guide A*. PMID: 24402721
46. Christofides NJ, Muirhead D, Jewkes RK, Penn-Kekana L, Conco DN. Women's experiences of and preferences for services after rape in South Africa: interview study. *BMJ*. 2006;332(7535):209. PMID: 16330476

47. Hanson K, McPake B, Nakamba P, Archard L. Preferences for hospital quality in Zambia: results from a discrete choice experiment. *Health Econ.* 2005;14(7):687-701. PMID: 15619273
48. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. *PharmacoEconomics.* 2019;37(2):201-26. PMID: 30392040
49. Janssen EM, Hauber AB, Bridges JFP. Conducting a Discrete-Choice Experiment Study Following Recommendations for Good Research Practices: An Application for Eliciting Patient Preferences for Diabetes Treatments. *Value in Health.* 2018;21(1):59-68. PMID: 29304942
50. Ferrini S, Scarpa R. Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management.* 2007;53(3):342-63. PMID:
51. Kessels R, Jones B, Goos P, Vandebroek M, L. . An Efficient Algorithm for Constructing Bayesian Optimal Choice Designs. KBI Working Paper No 0616, Available at SSRN: <https://ssrn.com/abstract=968620> or <http://dxdoi.org/102139/ssrn968620>. 2006. PMID:
52. Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling.* 2019;32:100170. PMID:
53. Johnson FR, Yang J-C, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. *Value in Health.* 2019;22(2):157-60. PMID:
54. Hess S, Daly A, Batley R. Revisiting consistency with random utility maximisation: theory and implications for practical work. *Theory and Decision.* 2018;84(2):181-204. PMID:
55. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ.* 2006;15(8):797-811. PMID: 16615039
56. Quaife M, Terris-Prestholt F, Eakle R, Cabrera Escobar MA, Kilbourne-Brook M, Mvundura M, et al. The cost-effectiveness of multi-purpose HIV and pregnancy prevention technologies in South Africa. *Journal of the International AIDS Society.* 2018;21(3):e25064. PMID: 29537654
57. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". *Translational Behavioral Medicine.* 2020;10(1):267-77. PMID:
58. Craig Lefebvre R, Tada Y, Hilfiker SW, Baur C. The Assessment of User Engagement with eHealth Content: The eHealth Engagement Scale<sup>1</sup>. *Journal of Computer-Mediated Communication.* 2010;15(4):666-81. PMID:
59. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology.* 2010;61(1):50-69. PMID:
60. Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy.* 2013;11(4):319-29. PMID: 23637054
61. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher 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. PMID: 23337210

62. Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 1991;50(2):179-211. PMID:
63. 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. *The European Journal of Health Economics*. 2018;19(8):1053-66. PMID: 29380229