Investigating the impact of design characteristics on statistical efficiency within discrete choice experiments: A systematic survey

Thuva Vanniyasingam[a,b,*], Caitlin Daly[a], Xuejing Jin[a], Yuan Zhang[a], Gary Foster[a,b], Charles Cunningham[c], Lehana Thabane[a,b,d,e,f]

[a] Department of Health Research Methods, Impact, and Evidence, McMaster University, Hamilton, ON, Canada
[b] Biostatistics Unit, Father Sean O'Sullivan Research Centre, St. Joseph's Healthcare, Hamilton, ON, Canada
[c] Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada
[d] Departments of Paediatrics and Anaesthesia, McMaster University, Hamilton, ON, Canada
[e] Centre for Evaluation of Medicine, St. Joseph's Healthcare, Hamilton, ON, Canada
[f] Population Health Research Institute, Hamilton Health Sciences, Hamilton, ON, Canada

ARTICLE INFO

ABSTRACT

Objectives: This study reviews simulation studies of discrete choice experiments to determine (i) how survey design features affect statistical efficiency, (ii) and to appraise their reporting quality.

Outcomes: Statistical efficiency was measured using relative design (D-) efficiency, D-optimality, or D-error.

Methods: For this systematic survey, we searched Journal Storage (JSTOR), Since Direct, PubMed, and OVID which included a search within EMBASE. Searches were conducted up to year 2016 for simulation studies investigating the impact of DCE design features on statistical efficiency. Studies were screened and data were extracted independently and in duplicate. Results for each included study were summarized by design characteristic. Previously developed criteria for reporting quality of simulation studies were also adapted and applied to each included study.

Results: Of 371 potentially relevant studies, 9 were found to be eligible, with several varying in study objectives. Statistical efficiency improved when increasing the number of choice tasks or alternatives; decreasing the number of attributes, attribute levels; using an unrestricted continuous "manipulator" attribute; using model-based approaches with covariates incorporating response behaviour; using sampling approaches that incorporate previous knowledge of response behaviour; incorporating heterogeneity in a model-based design; correctly specifying Bayesian priors; minimizing parameter prior variances; and using an appropriate method to create the DCE design for the research question. The simulation studies performed well in terms of reporting quality. Improvement is needed in regards to clearly specifying study objectives, number of failures, random number generators, starting seeds, and the software used.

Conclusion: These results identify the best approaches to structure a DCE. An investigator can manipulate design characteristics to help reduce response burden and increase statistical efficiency. Since studies varied in their objectives, conclusions were made on several design characteristics, however, the validity of each conclusion was limited. Further research should be conducted to explore all conclusions in various design settings and scenarios. Additional reviews to explore other statistical efficiency outcomes and databases can also be performed to enhance the conclusions identified from this review.

1. Introduction

Discrete choice experiments (DCEs) are now being used as a tool in health research to elicit participant preferences for a health product or service. Several DCEs have emerged within the health literature using various design approaches [3–6]. Ghijben and colleagues conducted a DCE to understand how patients value and trade-off key characteristics of oral anticoagulants [7]. They examined patient preferences to determine which of seven attributes of warfarin and other anticoagulants (dabigatran, rivaroxaban, apixaban) in atrial fibrillation were most important to patients [7]. With seven attributes, each with different levels, several possible combinations could be created to describe an anticoagulant. Like many DCEs, they used a fractional factorial design, a sample of all possible combinations, to create a survey with 16

---

questions that each presented three alternatives for patients to choose from. As patients selected their most preferred and second most preferred alternatives, investigators were able to model their responses to determine which anticoagulant attributes were more favourable than others. Since only a fraction of combinations are typically used in DCEs, it is important to use a statistical efficiency measure to determine how well the fraction represents all possible combinations of attributes and attribute levels.

There is no single specific design to yield optimal results of a discrete choice experiment (DCE). They can vary in their level of statistical efficiency and response burden. The variation in designs can be seen in several reviews covering various decades [8–13]. While presenting all possible combinations of attributes and attribute levels will always yield 100% statistical efficiency, this is not feasible in many cases. For fractional factorial designs, a statistical efficiency measure can be used to reduce the bias of the fraction selected. A common measure to assess statistical efficiency of these partial designs is relative design efficiency (D-efficiency) [14,15]. For a design matrix X, the formula is as follows:

$$Relative\ D - efficiency = 100 * \frac{1}{N_D\ |(X'X)^{-1}|^{1/p}},$$

where $X'X$ is the information matrix, p is the number of parameters, and N_D is the number of rows in the design [16].

To yield a statistically efficient design, a design will be orthogonal and balanced, or nearly so. A design is balanced when attribute levels are evenly distributed [17]. This occurs when the off-diagonal elements in the intercept row and column are zero [16]. It is orthogonal when the pairs of attribute levels are evenly distributed [17], that is when the submatrix *of X'X*, without the intercept, is a diagonal matrix [16]. Therefore, to maximize relative D-efficiency, we need to reduce the $(X'X)^{-1}$ matrix to a diagonal that equals $\frac{1}{N_d}I$ for a suitably coded X [16].

Relative D-efficiency is often referred to as relative design optimality (D-optimality) or is described using its inverse, design error (D-error) [18]. It ranges from 0% to 100%, where 100% indicates a statistically efficient design. A measure of 100% can still be achieved with fractional factorial designs; however, there is limited knowledge as to how the various design characteristics impact statistical efficiency.

Identifying the impact of DCE design characteristics on statistical efficiency will bring more power to investigators, particularly research practitioners, during the design stage. They can reduce the variance of estimates by manipulating their designs to construct a simpler DCE that is statistically efficient and minimizes participants' response burden. Currently there are several studies exploring DCE designs. These studies range from comparing or introducing new statistical optimality criteria [19,20] to approaches for generating DCEs [14] to exploring the impact of different prior specifications on parameter estimates [21–23]. To our knowledge, the results of these findings have not been summarized. This may be due to the variation in objectives and outcomes across studies, making it hard to synthesize information and draw conclusions. As part of a previous simulation study, a literature review was also performed to report the DCE design characteristics explored by investigators in simulation studies [1]. However, information on the impact of these design characteristics on relative D-efficiency, the common outcome among each study, was not assessed.

The primary aim of this systematic survey was to review simulation studies to determine design features that affect the statistical efficiency of DCEs—measured using relative D-efficiency, relative D-optimality, or D-error; and to appraise the completeness of reporting of the studies using the criteria for reporting simulation studies [24].

2. Methods

2.1. Eligibility criteria

The inclusion criteria were comprised of simulation studies of DCEs that explored the impact of DCE design characteristics on relative D-efficiency, D-optimality, or D-error. Search terms were first searched by variations in spelling and acronyms of individual terms and then combined into one search. Studies were restricted to English articles. Studies were excluded if they were not related to DCEs (or not referred to as stated preference, latent class, and conjoint analysis), were applications of DCEs, empirical comparisons, reviews or discussions of DCEs, or simulation studies that did not explore the impact of DCE design characteristics on statistical efficiency. Duplicate publications, meeting abstracts, letter, commentary, editorials and protocols, books and pamphlets were also excluded.

2.2. Search strategy

Two rounds of electronic searches were conducted covering the period from inception to Sept 19, 2016. The first round was performed on all databases from inception to July 20, 2016. The second round extended the search until Sept 19, 2016. The databases searched were Journal Storage (JSTOR), Science Direct, PubMed, and OVID which included a search within EMBASE. Studies identified from Vanniyasingam et al.'s literature review, that were not identified in this search, were also considered [1]. Table 1 (Supplementary Files) presents the detailed search strategy for each database.

2.3. Study selection

Four reviewers worked independently and in duplicate to screen titles and abstracts of all citations identified in the search. Any potentially eligible article identified by either reviewer, from each pair, proceeded to the full-text review stage. The same authors then, independently and in duplicate, applied the above eligibility criteria to screen the full text of these studies. Disagreement regarding eligibility were resolved through discussion. If a disagreement was unresolved, a third author (a statistician) adjudicated and resolved the conflict. After full-text screening forms were consolidated amongst pairs, data was extracted from eligible studies. Both the full-text screening and data extraction forms were first piloted with calibration exercises to ensure consistency in reviewer reporting.

2.4. Data extraction process

A Microsoft Excel spreadsheet was used to extract information related to general study characteristics, DCE design characteristics that varied or were held fixed, and the impact of the varied design characteristics on statistical efficiency.

2.5. Reporting quality

The quality of reporting was also assessed by extracting information related to the reporting guidelines for simulation studies described by Burton and colleagues [24]. Some components were modified to be more tailored for simulation studies of DCEs. This checklist included whether studies reported:

- A detailed protocol of all aspects of the simulation study
- Clearly defined aims

- The number of failures during the simulation process (defined as the number of times it was not possible to create a design given the design component restrictions)
- Software used to perform simulations
- The random number generator or starting seed, the method(s) for generating DCE datasets
- The scenarios to be investigated (defined as the specifications of each design characteristic explored and overall total number of designs created)
- Methods for evaluating each scenario
- The distribution used to simulate the data (defined as whether or not the design characteristics explored are motivated by real-world or simulation studies)
- The presentation of the simulation results (defined as whether authors used separate subheadings for objectives, methods, results and discussion to assist with clarity). Information presented in graphs or tabular form but not written as detailed in the manuscripts were counted for if they were presented in a clear and concise manner.

One item was added to the criteria to determine whether or not studies provided a rationale for creating the different designs. Reporting items excluded were: a detailed protocol of all aspects of the simulation study, level of dependence between simulated designs, estimates to be stored for each simulation, summary measures to be calculated over all simulations, and criteria to evaluate the performance of statistical methods (bias, accuracy, and coverage). We decided against checking whether a detailed protocol was reported because the studies of interest were focussed on only the creation of DCE designs. The original reporting checklist is tailored towards randomized controlled trials or prognostic factor studies with complex situations seen in practice [24]. The remaining items were excluded because the specific statistical efficiency measures were required for studies to be included in the study. Also, there were no summary measures to be calculated over all simulations, and no results to measure bias, accuracy and coverage.

When studies referred to supplementary materials, these materials were also reviewed for data extraction.

Three of the four reviewers, working in pairs, performed data abstraction independently and in duplicate. Pairs resolved disagreements through discussion or, if necessary, with assistance from another statistician.

3. Data analysis

3.1. Agreement

Agreement between reviewers on the studies' eligibility based on full text screening was assessed using an unweighted kappa. A kappa value was indicative of poor agreement if it was less than 0.00, slight agreement if it ranged from 0.00 to 0.20, fair agreement between 0.21 and 0.40, moderate agreement between 0.41 and 0.60, substantial agreement between 0.61 and 0.80, and almost perfect agreement when greater than 0.80 [25].

3.2. Data synthesis and analysis

The simulation studies were assessed by the details of their DCE designs. More specifically, the design characteristics investigated and their ranges were recorded along with their impact on statistical efficiency (relative D-efficiency, D-optimality, or D-error). Adherence to reporting guidelines was also recorded [24].

4. Results

4.1. Search strategy and screening

A total of 371 papers were identified from the search and six were selected from a previous literature search that used snowball sampling [1]. From this, 43 were removed as duplicates and 245 were excluded during title and abstract screening. Of the remaining 77 studies for full text screening, three needed to be ordered [26–28] and one we were unable to obtain a full text for [29], 18 did not relate to DCEs (or include terms such as discrete choice, DCE, choice-based, binary choice, stated preference, latent class, conjoint analysis, or fractional factorial design, factorial design); 17 did not perform a simulation analysis; 1 did not use its simulations to create DCE designs; 22 did not assess the statistical efficiency of designs using relative D-optimality, D-efficiency, or D-error measures; 4 did not compare the impact of various design characteristics on relative D-efficiency or D-optimality or D-error; and 1 was not a peer-reviewed manuscript. Details of the search and screening process are presented in a flow chart in Fig. 1(Appendix).

Finally, nine studies remained after full-text screening. The unweighted kappa for measuring agreement between reviewers on full text eligibility was 0.53, indicating a moderate agreement [25]. Of the 9 studies included, 1 was published in Marketing Science, 1 in the Journal of Statistical Planning and Inference, 2 in the Journal of Marketing Research, 1 in the International Journal of Research in Marketing, 2 in Computational Statistics and Data Analysis, 1 in BMJ Open, and 1 in Transportation Research Part B: Methodological.

The number of statistical efficiency measures, scenarios, and design characteristics varied from study to study. Of the outcomes assessed for each scenario, four studies reported relative D-efficiency [1,30–32], two D-error [33,34], three D_b-error (a Bayesian variation of D-error) [30,35,36], and two percentage changes in D-error [34,37]. Of the design characteristics explored, one study explored the impact of attributes on statistical efficiency [1], two explored alternatives [1,30], one explored choice tasks [1], two explored attribute levels [1,32], two explored choice behaviour [33,37], three explored priors [30,31,34], and four explored methods to create the design [30,34–36]. Results are further described below based for each design characteristic. Details of the ranges of each design characteristic investigated and corresponding studies are described in Table 2 (Supplementary Files).

4.2. Survey-specific components

The simulation studies had several conclusions based on the number of choice tasks, attributes, and attribute levels; the type of attributes (qualitative and quantitative); and the number of alternatives. First, ***increasing the number of choice tasks*** increased relative D-efficiency (or improved statistical efficiency) across several designs with varying numbers of attributes, attribute levels, and alternatives [1]. Second, ***increasing the number of attributes*** generally (i.e. not monotonically) decreased relative D-efficiency. For designs with a large number of attributes and a small number of alternatives per choice task, a DCE could not be created [1]. Third, ***increasing the number of levels within attributes*** (from 2 to 5) decreased relative D-efficiency. In fact, binary attribute designs had higher statistical efficiency in comparison to all other designs with varying numbers of alternatives (2–5), attributes (2–20), and choice tasks (2–20). However, higher relative D-efficiency measures were also found when the number of attribute levels equalled the number of alternative [1]. Fourth, ***increasing the number of alternatives*** improved statistical efficiency [1,30]. Fifth, for ***designs with only binary attributes and one quantitative (continuous) attribute, it***

was possible to create locally optimal designs. To further clarify this result, DCEs were created where two of three alternatives were identical or differed only by an unrestricted continuous attribute (e.g. *size, weight, or speed*). The third alternative differed from the two others in the binary attributes [32]. The continuous variable was unrestricted and used as a "manipulating" attribute to offset dominating alternatives or alternatives with a zero probability of being selected in a choice task. This finding, however, was conditional on the type of quantitative variable and was concluded to be unrealistic in the study [32]. Details of the studies exploring these design characteristics are presented in Tables 1a and 1b (Appendix).

4.3. Incorporating choice behaviour

Two approaches were used to incorporate response behaviour when designing a DCE. First, the order of the statistical efficiency of designs from highest to lowest were if they: (i) incorporated covariates relating to response behaviour, (ii) incorporated covariates not relating to response behaviour, and (iii) did not incorporate any covariates. Second, among binary choice designs, stratified sampling strategies had higher statistical efficiency measures in comparison to randomly sampled strategies. This was most apparent when stratification was performed on both expected choice behaviour (e.g. 2.5% of the population selects $Y = 1$, remaining selects $Y = 0$) and on a binary independent factor associated with the response behaviour. Similar efficiency measures were found when there was an even distribution (50% of the population selects $Y = 1$) across approaches [37] (Table 1c, Appendix).

4.4. Bayesian priors

Studies also explored the impact of parameter priors and heterogeneity priors. Increasing the parameter prior variances [30] or misspecifying priors (in comparison to correctly specifying priors) [34] reduced statistical efficiency. In one study, mixed logit designs that incorporated respondent heterogeneity had higher statistical efficiency measures than designs ignoring respondent heterogeneity [31]. However, misspecifying the heterogeneity prior had negative implications. In fact, underspecifying the heterogeneity prior had a greater loss in efficiency in comparison to over specifying it [31] (Table 1d, Appendix).

4.5. Methods to create the design

Several simulation studies compared various methods to create a DCE design against other design settings (Table 1e, Appendix). First, relative statistical efficiency measures were highest when the method to create a design matched the method used for the reference design setting [30,34,36]. For example, a multinomial logit (MNL) generated design had the highest statistical efficiency in an MNL design setting, in comparison to a cross-sectional mixed logit or a panel mixed logit design setting [36]. Similarly, a partial rank-order conjoint experiment yields highest statistical efficiency for a design setting of the same type in comparison to a best-choice experiment, best-worst experiment or orthogonal design setting [30]. Second, among frequentist (non-Bayesian) approaches, the order of designs yielding the highest statistical efficiency is d-optimal rank designs, d-optimal choice designs, near-orthogonal, random designs, and balanced overlap designs for full rank order and partial rank order choice experiments [35]. Third, a semi-Bayesian d-optimal best-worst choice design outperformed frequentist and Bayesian-derived designs, while yielding similar statistical efficiency measures as semi-Bayesian d-optimal best-worst choice designs [30].

4.6. Reporting of simulations studies

All studies clearly reported the primary outcome, rationale and methods for creating designs, and methods to evaluate each scenario. Reporting the objective was unclear in two studies and no study reported any failures in the simulations. In many cases, such as in Vermeulen et al.'s study [36], the distribution from which random numbers were selected from were described, however no study specified the starting seeds. Also, no study reported the number of times it was not possible to create a design given the design component restrictions except for Vanniyasingam et al. [1], who specified that designs with a larger number of attributes could not be created with a small number of alternatives or choice task. The total number of designs and the range of design characteristics explored were either written or easily identifiable from figures and tables. Five studies reported the software used for the simulation studies and one study reported the software used for only one of the approaches to create a design. Four studies chose design characteristics that were motivated by real-world scenarios or previous literature, while four were not motivated by other studies. Details of each study's reporting quality are broken down in Table 2 (Appendix).

5. Discussion

5.1. Summary of findings

Several conclusions can be drawn from the nine simulation studies included in this systematic survey of investigating the impact of design characteristics on statistical efficiency. Factors recognized for improving statistical efficiency of a DCE include (i) increasing the number of choice tasks or alternatives; (ii) decreasing the number of attributes, and levels within attributes; (iii) using model-based designs with covariates or sampling approaches that incorporate response behaviour; (iv) incorporating heterogeneity in a model-based design; (v) correctly specifying Bayesian priors and minimizing parameter prior variances; and (vi) the method to create the DCE design is appropriate for the research question and design at hand. Lastly, optimal designs could be created using 3 alternatives with all binary attributes except one continuous attribute. Here, two alternatives were identical or differed only by the continuous attribute and the third alternative differed by the binary attribute. Overall, studies were detailed in their descriptions of simulation studies. Improvement is needed to ensure the study objectives, number of failures, random number generators, starting seeds, and the software used are clearly defined.

5.2. Discussion of simulation studies

Many of the studies agree with the formula for relative d-efficiency, however some appear to contradict it. Conclusions related to choice tasks, alternatives, attributes, and attribute levels all agree with the relative d-efficiency formula where increasing the number of parameters (with attributes and attribute levels) will reduce statistical efficiency and increasing the number of choice tasks improves it. Also, when the number of attribute levels and alternatives are equal, increasing the number of attribute levels may compromise statistical efficiency, however it can be compensated by increasing the number of alternatives (which may increase N_d). A conclusion that cannot be directly deduced from the formula are in relation to designs with qualitative and unrestricted quantitative attributes. Grabhoff and colleagues were able to create optimal designs where two alternatives were either completely identical or only differed by a continuous variable [32]. With less information provided within each choice task (or more

overlaps), we expect a lower statistical efficiency measure. Their design approach first develops a design solution using the binary attributes and then adds the continuous attribute to maximize the efficiency. This was a continuation of Kanninen's study who explained that the continuous attribute could be used to offset dominating alternatives or alternatives that carried a zero probability of being selected by a respondent [38]. It acted as a function of a linear combination of the other binary attributes. This continuous attribute, however, was conditional on the type of quantitative variable (such as size). Other types (such as price) may result in the "red bus/blue bus" parody) [32].

5.2.1. Importance

To our knowledge, this systematic survey is the first of its kind in synthesizing information on the impact of DCE design characteristics on statistical efficiency in simulation studies. Other studies have focussed on the reporting of applications of DCEs [39], and the details of DCEs and alternative approaches [40]. Systematic and literature reviews have highlighted the design type (e.g. fractional factorial or full factorial designs) and statistical methods used to analyze applications of DCEs within health research [2,11,13,41]. Exploration into summarizing the results of simulation studies is limited.

5.2.2. Strengths

This study has several strengths. First, it focuses on simulation studies which are able to (i) explore several design settings to answer a research question in a single study that real world applications are unable to; (ii) act as an instrumental tool to aid in the understanding of statistical concepts such as relative d-efficiency; and identify patterns in design characteristics for improving statistical efficiency. Second, it appraises the rigour of the simulations performed, through evaluating the reporting quality, to ensure the selected studies are appropriately reflecting high quality DCEs. Third, it provides an overview for investigators to assess the scope of the literature for future simulation studies. Fourth, the results presented here can provide further insight for investigators on patterns that exist in statistical efficiency. For example, if some design characteristics must be fixed (such as the number of attributes and attribute levels), investigators can manipulate others (e.g. number of alternatives or choice tasks) to improve both the statistical optimality and response efficiency of the DCE.

5.2.3. Limitations

There are some caveats to this systematic survey that may limit the direct transferability of these results to empirical research. First, the search for simulation studies of DCEs was only performed within health databases. Despite capturing a few studies from marketing journals in our search, we did not explore grey literature, statistics journals, or marketing journals. Second, we only describe the results for three outcomes (relative D-efficiency, D-error, and D-optimality) while some studies have reported other statistical efficiency measures. Third, with only nine included studies, each varying in objectives, it was not possible to make strong conclusions at this stage. Only summary findings of each study could be presented. Last, informant (or response) efficiency was not considered when extracting results from each simulation study. We recognize that incorporating participants' cognitive burden has a critical impact on the effect of overall estimation precision[42]. Integrating response efficiency with statistical efficiency would refine the focus on the structure, content, and pretesting of the survey instrument itself.

5.2.4. Further research

This systematic survey provides many avenues for further research. First, these results can be used as hypotheses for future simulation studies to test and compare in various DCE scenarios. Second, a review can be performed on other statistical efficiency outcomes such as the precision of parameter estimates or reduction in sample size to compare the impact of each design characteristic. Third, a larger review should be conducted to explore simulation studies within economic, marketing, and pharmacoeconomic databases.

6. Conclusions

Presenting as many possible combinations (via choice tasks or alternatives) or decreasing the total number of all possible combinations (via attributes or attribute levels) will improve statistical efficiency. Model-based approaches were popularly used to create designs. These models varied from adjusting for heterogeneity, including covariates, and using a Bayesian approach. They were also applied to several different design settings. Overall reporting was clear, however improvements can be made to ensure the study objectives, number of failures, random number generators, starting seeds, and the software used are clearly defined. Further areas of research to aid in solidifying the conclusions from this paper include a systematic survey of other outcomes related to statistical efficiency, a survey on databases outside of health research that also use DCEs, and a large-scale simulation study to test each conclusion from these simulation studies.

Competing interests

TV, CD, XJ, YZ, GF, and LT declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. CC's participation was supported by the Jack Laidlaw Chair in Patient-Centered Health Care.

Data sharing

As this is a systematic survey of simulation studies no data of patients exist. Data from each simulation study needs to be obtained directly from their corresponding authors.

Author contributions

All authors provided intellectual content for the manuscript and approved the final draft.

TV contributed to the conception and design of the study; screened and extracted data; performed the statistical analyses; drafted the manuscript; approved the final manuscript; and agrees to be accountable for all aspects of the work in relation to accuracy or integrity.

CD, XJ, and YJ assisted in screening and extracting data; critically assessed the manuscript for important intellectual content; approved the final manuscript; and agrees to be accountable for all aspects of the work in relation to accuracy or integrity.

LT contributed to the conception and design of the study; provided statistical and methodological support in interpreting results and drafting the manuscript; approved the final manuscript; and agrees to be accountable for all aspects of the work in relation to accuracy or integrity.

CC and GF contributed to the design of the study; critically assessed the manuscript for important intellectual content; approved the final manuscript; and agree to be accountable for all aspects of the work in relation to accuracy or integrity.
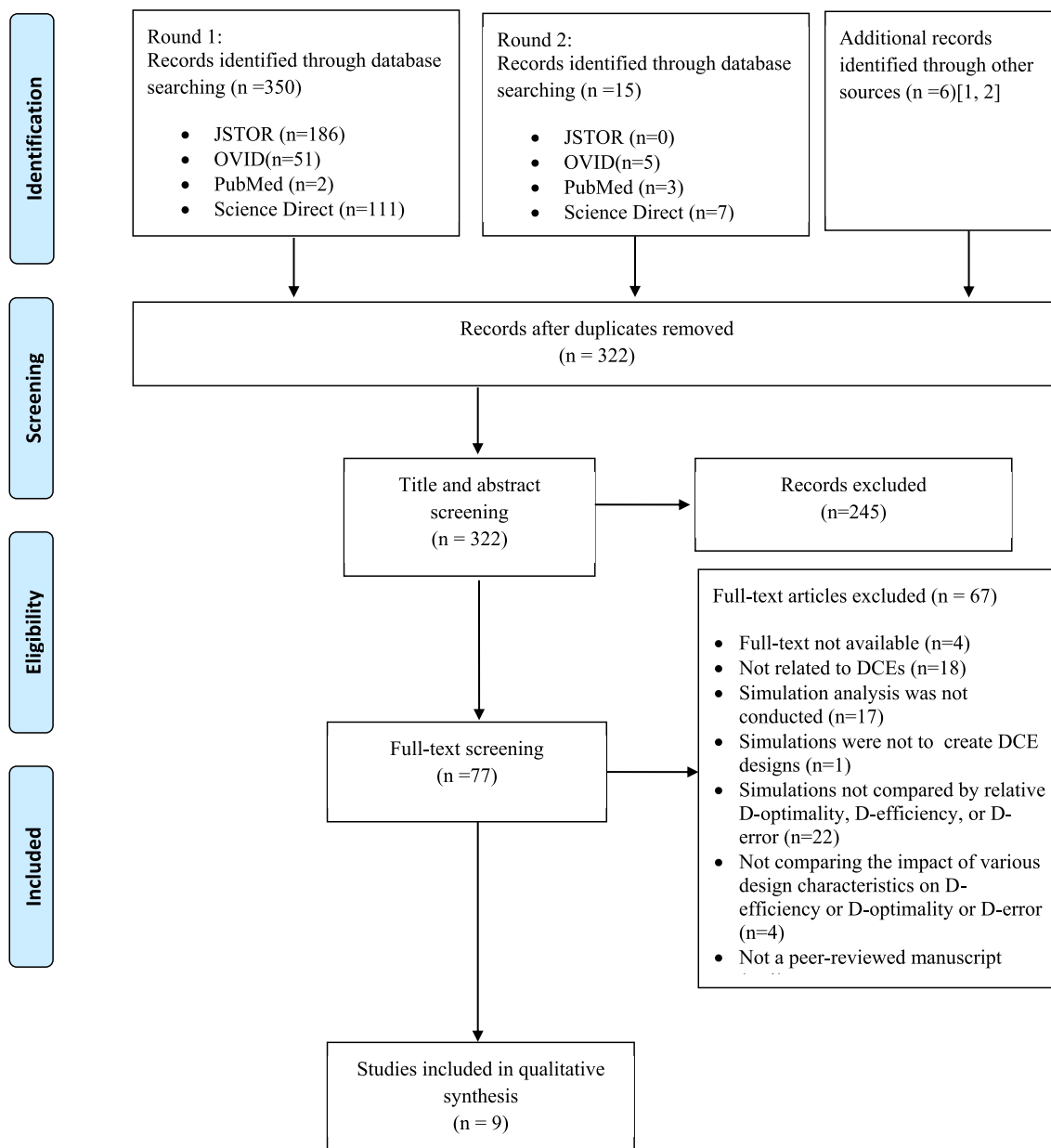
Appendix

Identification

> Round 1:
> Records identified through database searching (n =350)
>
> - JSTOR (n=186)
> - OVID(n=51)
> - PubMed (n=2)
> - Science Direct (n=111)

> Round 2:
> Records identified through database searching (n =15)
>
> - JSTOR (n=0)
> - OVID(n=5)
> - PubMed (n=3)
> - Science Direct (n=7)

> Additional records identified through other sources (n =6)[1, 2]

Screening

> Records after duplicates removed
> (n = 322)

> Title and abstract screening
> (n = 322)

> Records excluded
> (n=245)

Eligibility

> Full-text articles excluded (n = 67)
>
> - Full-text not available (n=4)
> - Not related to DCEs (n=18)
> - Simulation analysis was not conducted (n=17)
> - Simulations were not to create DCE designs (n=1)
> - Simulations not compared by relative D-optimality, D-efficiency, or D-error (n=22)
> - Not comparing the impact of various design characteristics on D-efficiency or D-optimality or D-error (n=4)
> - Not a peer-reviewed manuscript

Included

> Full-text screening
> (n =77)

> Studies included in qualitative synthesis
> (n = 9)

Fig. 1. PRISMA flow diagram.

Table 1a
Studies investigating the number of choice tasks, attributes, and attribute levels

Author, Year	Outcome of interest	Method to create design	Design setting	Distribution of Priors of parameter estimates	Choice sets	Alternatives	Attributes	Attribute levels	Results
# Choice tasks									
Vanniyasingam, 2016 [1]	Relative D-efficiency	.	.	no priors	2–20	2–5	2–20	2–5	1) Generally, as the number of choice tasks increases, relative D-efficiency increases
# Attributes									
Vanniyasingam, 2016 [1]	Relative D-efficiency	.	.	no priors	2–20	2–5	2–20	2–5	1) Generally, increasing # attributes, decreases relative D-efficiency (not monotonically) 2) designs with a small # of alternatives and large number of attributes could not be created.
# Attribute levels									
Vanniyasingam, 2016 [1]	Relative D-efficiency	.	.	no priors	2–20	2–5	2–20	2–5	1) Generally, increasing # attribute levels, decreases relative D-efficiency 2) designs yield higher D-efficiency measures when the # attribute levels match the number of alternatives 3) Generally, binary attributes perform best across all other designs
Graβhoff, 2013 [32]	Efficiency	.	.	$\beta_1 = 0$, $\beta_2 = 1$	.	3	1–7	Unrestricted quantitative (continuous) and qualitative (binary) attributes	1) Design-optimality was achieved when two alternatives were identical or differed only in the (unrestricted) quantitative variable, while the third alternative varies in all of the qualitative components.

Table 1b
Studies investigating the number of alternatives on statistical efficiency

Author, Year	Outcome of interest	Method to create design	Design setting	Distribution of Priors of parameter estimates	Choice sets	Alternatives	Attributes	Attribute levels	Results
# Alternatives									
Vermeluen, 2010 [30]	D_b-error	1: Best choice experiment 2: Partial rank-order conjoint experiment 3: Best-worst choice experiment	1: Partial best choice experiment 2: Rank-order conjoint experiment 3: Best-worst experiment	Parameter estimates follow a normal distribution with mean priors [-0.75 0–0.75 0–0.75 0 0.75 0.75] Variance priors range: a) 0.04 b) 0.5	9	4, 5, 6	5	3 three-level, 2 two-level attributes	Alternatives 1) Increasing the number of alternatives reduced the D_b-error across all scenarios

| Vanniyasingam, 2016 [1] | Relative D-efficiency | 4: Orthogonal designs; Random allocation | No priors | 2–20 | 2–5 | 2–20 | 2–5 | 2–20 | 2–5 | Alternatives: 1) Increasing # alternatives, increases relative D-efficiency except for binary attributes which best performed with only 2 alternatives 2) designs yield higher D-efficiency measures when the # attribute levels match the number of alternatives |

Table 1c
Studies investigating the incorporation of choice behaviour on statistical efficiency

Author, Year	Outcome of interest	Method to create design	Design setting	Sample size	Choice sets	Alternatives	Attributes	Attribute levels	Results
Crabbe, 2012 [33]	Local D-error	1) Individually adapted sequential Bayesian designs (IASB) with covariates incorporated 2) IASB designs, no covariates 3) single nearly orthogonal designs, no covariates	1) Choice behaviour is influenced by 2 covariates 2) Choice behaviour is NOT influenced by 2 (irrelevant) covariates	25, 250	16	3	3	3	Across all design settings and sample sizes: 1) Despite IASB designs incorporating two *irrelevant* covariates of participant behaviour, they still more statistically efficient than designs that do not incorporate any covariates. 2) IASB designs with two relevant covariates perform better (in terms of D-efficiency) in comparison to IASB designs with to irrelevant covariates, holding everything else constant.
Donkers, 2003 [37]	Average percentage change in D-error	Design incorporates the proportion of the population selecting y = 1, which varies from 2.5%, 5%, 10%, 15%, and 50% of the population. Results of D-error compared to random sampling from population.	.	Sample selection is dependent on proportion that selects Y = 1		2	2	1 binary, 1 continuous (Distribution of binary attribute when X = 1: 50% or 10% of the time)	1) As the proportion of the population selecting Y = 1 increases from 2.5% to 50%, D-efficiency improves. 2) As more individuals select 1, the magnitude of the reduction in D-error decreases (in comparison to when a random sample is used). The highest reduction in D-error (or improvement in D-efficiency) is when only 2.5% of the population selects y = 1. 3) Above results are consistent when binary attribute (x = 1) is distributed 10% or 50% of the time within the DCE.
Donkers, 2003 [37]	Average percentage change in D-error	Design incorporates the proportion of the population selecting y = 1, which varies from 2.5%, 5%, 10%, 15%, and 50% of the population. Results of D-error compared to random sampling from population.	.	Sample selection is dependent on: a) y only b) y and x c) x only		2	2	1 binary, 1 continuous; binary attribute is unevenly distributed with x = 1 only 10% of the time	Type of sample selection (y only, y and x, x only) 1) Designs with sample selection on both y and x yields higher statistical efficiency than designs with sample selection on y only or x only, where y is the outcome, and x is an attribute.

Table 1d

Studies investigating Bayesian priors on statistical efficiency

Author, Year	Yu, 2009 [31]	Vermeulen, 2010 [30]	Bliemer, 2010 [34]
Outcome of interest	Relative local D-efficiency	D_b-error	D-error and percentage change in D-error
Describe the scenario	8 different designs, each compared within 5 different parameter spaces/design settings.	Comparing four designs within 3 settings for designs varying in alternatives and variance priors of parameters	Misspecification of prior parameter values
Method to create design	Models 1–3: Mixed logit semi-Bayesian d-optimal design Model 4: ML locally d-optimal design Models 5,6: MNL Bayesian D-optimal design Model 7: MNL Locally D-optimal design Model 8: Nearly orthogonal design	Model 1: Best choice experiment Model 2: Partial rank-order conjoint experiment Model 3: Best-worst choice experiment Model 4: Orthogonal designs	Model 1: MNL Model 2: Cross-sectional mixed logit Model 3: Panel mixed logit
Design setting	Parameters were drawn from a normal distribution: Mean: Models 1–3 = μ; Model 4: μ; Model 5: μ+ 0.51 × I_8; where μ = [-0.5 0-0.5 0-0.5 0-0.5 0]' Covariance: Model 1 = 0.25 × I_8; Model 2 = I_8; Model 3 = 2.25 × I_8; Model 4 = I_8; Model 5 = I_8	Setting 1: Partial best choice experiment Setting 2: Rank-order conjoint experiment Setting 3: Best-worst experiment	Setting 1: MNL model Setting 2: Cross-sectional mixed logit model Setting 3: Panel mixed logit model
Heterogeneity prior	Model 1 = 1.5 × 18; Model 2 = 1_8; Model 3: 0.5 × 18; Model 4 = 1_8; Model 5–7: 0_8; Model 8: no prior, Where 1_8 is an 8-dimensional identity matrix	.	.
Distribution of Priors of parameter estimates	Model 1–3: Normal distribution with fixed mean, covariance I8 Model 4: fixed mean Model 5: Normal distribution with fixed mean, covariance 9 × 18 Model 6: Normal distribution with fixed mean, covariance I8 Model 7: fixed mean Model 8: no priors Where fixed mean vector is: [-0.5 0-0.5 0-0.5 0-0.5 0]'	Parameter estimates follow a normal distribution with mean priors: [-0.75 0–0.75 0–0.75 0 0.75 0.75], and variance prior range: a) 0.04 b) 0.5	Settings 1–3: Assumed true value of parameters: β0 = − 0.5, β1: Normal (−0.07, 0.03), β2: Uniform (−1.1, − 0.8), β3: Normal (−0.6, 0.15), β4 = −0.3 Designs 1–3 (from scenario 3): β0 = − 0.5, β1: Normal (− 0.05, 0.02), β2: Uniform (−0.9, 0.2), β3: Normal (− 0.8, 0.2), β4 = − 0.2
Choice sets	12	9	12
Alternatives	3	4, 5, 6	2
Attributes	4	5	3
Attribute levels	3	3 three-level, 2 two-level attributes	2 three-level attributes, 1 four-level attribute
Results	1) Across all 5 design settings: Mixed logit model designs performed substantially better than designs that ignored respondent heterogeneity 2) Comparing Semi-Bayesian designs (Models 1–3): a) Overspecifying the heterogeneity prior (Model 1) does not have too large of a negative impact on efficiency b) Underspecifying the heterogeneity prior (Model 3) has a greater loss in efficiency in comparison to overspecifying it (Model 1) 3) Results remain consistent across other design setting such as: 2 × 3 × 4/2/24 and 2 × 2 × 3/3/12	Parameter priors: 1) Increasing the parameter prior variances increased the d_b-error.	1) D-errors of designs with misspecified priors were higher than designs with correctly specified priors (from scenario.

Table 1e

Studies investigating methods to create DCE designs on statistical efficiency

Author, Year	Vermeulen, 2011 [35]	Bliemer, 2010 [34]	Vermeulen, 2008 [36]	Vermeulen, 2010 [30]	Vermeulen, 2010 [30]
Outcome	D_b-error	D-error	d_b-error	D_b-error	Relative D-efficiencies
Describe the scenario	Comparing different designs to create DCEs for 2 settings: full rank- and partial rank-order choice-based conjoint experiments	Comparing three types of designs against each other and an orthogonal design.	Comparing different designs to create DCEs in 2 settings: a presence and absence of 'no-choice' alternative in DCEs	Comparing different designs to create DCEs in 3 settings and with varying alternatives and variance priors	Comparing semi-Bayesian D-optimal best-worst design with 6 benchmark designs
Choice sets	9	9, 12	16	9	9
Alternatives	4	2,3	2 and "no choice" alternative	4,5, 6	4
Attributes	5	3,4	3	5	5
Attribute levels	$3^3 2^2$	$3^2 4^1$	$3^2 2^1$	$3^2 2^2$	$3^2 2^2$
Method to create design	Design: 1. Bayesian D-optimal ranking 2. D-optimal choice 3. Balanced overlap 4. Near-orthogonal 5. Random	Design: 1. MNL design 2. Cross-sectional mixed logit design (heterogeneity prior = 0), Panel mixed logit design Priors: fixed parameters, priors equal to the mean	Design: 1. MNL model 2. Extended no-choice MNL 3. Nested no-choice MNL 4. Model-robust	Design: 1. Best choice 2. Partial rank-order conjoint 3. Best-worst choice 4. Orthogonal	Design: 1. Semi-Bayesian D-optimal best-worst choice 2. Utility-neutral best-worst choice 3. Semi-Bayesian D-optimal choice 4. Utility-neutral choice 5. Nearly orthogonal 6. Random 7. Balanced attribute level overlap
Design setting	Setting: 1. Full rank-order choice-based conjoint experiments 2. Partial rank-order choice-based conjoint experiments	Setting: 1. MNL 2. Cross-sectional mixed logit 3. Panel mixed logit model 4. Orthogonal (within alternatives) design	Setting: 1. Extended no-choice multinomial logit model 2. Nested no-choice multinomial logit model	Setting: 1. Partial best choice experiment 2. Rank-order conjoint experiment 3. Best-worst experiment	Setting: Design 1 was set as the comparator design against all other designs (above#2–7)
Priors		Settings 1–3: Assumed priors correspond to true parameter values: Case 1: $\beta_1 \sim N(0.6, 0.2)$, $\beta_2 \sim N(-0.9, 0.2)$, $\beta_3 = -0.2$, $\beta_4 = 0.8$; Case 2: $\beta_1 \sim U(-0.9, 0–0.5)$, $\beta_2 \sim N(-0.8, 0.2)$, $\beta_3 \sim U(-1.5, -1.0)$		Priors for each setting: Parameter estimates follow a normal distribution with mean priors: [-0.75 0–0.75 0–0.75 0 0.75 0.75] and variance priors: a) 0.04 b) 0.5	Coefficients come from an 8-dimensional normal distribution with mean prior: [1.5 0 1.5 0 1.5 1.5 1.5] and variance prior

Results		(for every coefficient): 0.5.
D-opt rank > D-opt. choice > Near-orthogonal > Random > Balanced overlap	Models estimated using designs specifically generated for that model outperform designs generated for different mode forms.	Case 3: $\beta0 = -0.5$, $\beta_1 \sim N(-0.05, 0.02)$, $\beta_2 \sim U (-0.9, 0.2)$, $\beta_3 \sim N(-0.8, 0.2)$, $\beta_4 = -0.2$; Setting 4: Misspecification of prior parameters
	For setting 1: Model 2 > 4 > 3 > 1	Models estimated using designs specifically generated for that model outperform designs generated for different mode forms.
	For setting 2: Model 3 > 4 > 2 > 1	1)Models estimated using designs specifically generated for that model outperform designs generated for different mode forms. 2) Models 1,2,3 > 4
		1. Design 1 > 2, 4, 5, 6, 7 2. Design 1's optimality is similar to Design 3 outstanding

Comment: The greater than sign " > " indicates which method performed better than another method in terms of statistical efficiency.

Table 2
Reporting items of simulations studies

Author, Year	Protocol	Primary outcome	Clear aim	Number of failures	Software	Random number generator or starting seed	Rationale for creating designs	Methods for creating designs	Scenarios; Total number of designs	Scenarios: Range of design characteristics explored	Method to evaluate each scenario	Distribution used to simulate data*
Vermeulen, 2011 [35]	0	1	1	0	1	0	1	1	1	1	1	0
Yu, 2009 [31]	0	1	1	0	0, 1	0	1	1	1	1	1	1
Bliemer, 2010 [34]	0	1	0	0	1	0	1	1	1	1	1	1
Crabbe, 2012 [33]	0	1	1	0	1	0	1	1	1	1	1	0
Vermeulen, 2010 [30]	0	1	1	0	1	0	1	1	1	1	1	0
Vermeulen, 2008 [36]	0	1	1	0	0	0	1	1	1	1	1	0
Vanniyasingam, 2016 [1]	0	1	1	1	1	0	1	1	1	1	1	1
Graßhoff, 2013 [32]	0	1	0	0	0	0	1	1	1	1	1	1
Donkers, 2003 [37]	0	1	1	0	0	0	1	1	1	1	1	0

Comment: 1 = reported; 0 = unclear/not reported for each column.
*1 = the chosen design characteristics are motivated by real-world scenario (previous literature referenced, etc) OR by other simulation study scenarios, 0 = not motivated by other studies.

Appendix B. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.conctc.2018.01.002.

References

[1] T. Vanniyasingam, C.E. Cunningham, G. Foster, L. Thabane, Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments, BMJ open 6 (7) (2016) 2016–011985.

[2] M.D. Clark, D. Determann, S. Petrou, D. Moro, E.W. de Bekker-Grob, Discrete choice experiments in health economics: a review of the literature, Pharmacoeconomics 32 (9) (2014) 883–902.

[3] V.H. Decalf, A.M.J. Huion, D.F. Benoit, M.A. Denys, M. Petrovic, K. Everaert, Older People's preferences for side effects associated with antimuscarinic treatments of overactive bladder: a discrete-choice experiment, Drugs Aging (2017).

[4] J.M. Gonzalez, S. Ogale, R. Morlock, J. Posner, B. Hauber, N. Sommer, A. Grothey, Patient and physician preferences for anticancer drugs for the treatment of metastatic colorectal cancer: a discrete-choice experiment, Canc. Manag. Res. 9 (2017) 149–158.

[5] M. Feehan, M. Walsh, J. Godin, D. Sundwall, M.A. Munger, Patient preferences for healthcare delivery through community pharmacy settings in the USA: a discrete choice study, J. Clin. Pharm. Therapeut. (2017).

[6] A. Liede, C.A. Mansfield, K.A. Metcalfe, M.A. Price, C. Snyder, H.T. Lynch, S. Friedman, J. Amelio, J. Posner, S.A. Narod, et al., Preferences for breast cancer risk reduction among BRCA1/BRCA2 mutation carriers: a discrete-choice experiment, Breast Canc. Res. Treat. (2017).

[7] P. Ghijben, E. Lancsar, S. Zavarsek, Preferences for oral anticoagulants in atrial fibrillation: a best–best discrete choice experiment, Pharmacoeconomics 32 (11) (2014) 1115–1127.

[8] M. Ryan, K. Gerard, Using discrete choice experiments to value health care programmes: current practice and future research reflections, Appl. Health Econ. Health Pol. 2 (1) (2003) 55–64.

[9] M. Lagarde, D. Blaauw, A review of the application and contribution of discrete choice experiments to inform human resources policy interventions, Hum. Resour. Health 7 (1) (2009) 1.

[10] M.C. Bliemer, J.M. Rose, Experimental design influences on stated choice outputs: an empirical study in air travel choice, Transport. Res. Pol. Pract. 45 (1) (2011) 63–79.

[11] E.W. de Bekker-Grob, M. Ryan, K. Gerard, Discrete choice experiments in health economics: a review of the literature, Health Econ. 21 (2) (2012) 145–172.

[12] D. Marshall, J.F. Bridges, B. Hauber, R. Cameron, L. Donnalley, K. Fyie, F.R. Johnson, Conjoint analysis applications in health—how are studies being designed and reported? Patient Patient-Cent. Outcomes Res. 3 (4) (2010) 249–256.

[13] K.L. Mandeville, M. Lagarde, K. Hanson, The use of discrete choice experiments to inform health workforce policy: a systematic review, BMC Health Serv. Res. 14 (1) (2014) 1.

[14] L. Lourenço-Gomes, L.M.C. Pinto, J. Rebelo, Using choice experiments to value a world cultural heritage site: reflections on the experimental design, J. Appl. Econ. 16 (2) (2013) 303–332.

[15] J.J. Louviere, D. Street, L. Burgess, N. Wasi, T. Islam, A.A.J. Marley, Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information, J. Choice Modell. 1 (1) (2008) 128–164.

[16] W.F. Kuhfeld, Experimental design, efficiency, coding, and choice designs, Marketing Research Methods in Sas: Experimental Design, Choice, Conjoint, and Graphical Techniques, Iowa State University: SAS Institute Inc., 2005, pp. 53–241.

[17] W.F. Kuhfeld, Marketing Research Methods in SAS. Experimental Design, Choice, Conjoint, and Graphical Techniques Cary, NC, SAS-Institute TS-722, 2005.

[18] K. Zwerina, J. Huber, W.F. Kuhfeld, A General Method for Constructing Efficient Choice Designs, Fuqua School of Business, Duke University, Durham, NC, 1996.

[19] O. Toubia, J.R. Hauser, On managerially efficient experimental designs, Market. Sci. 26 (6) (2007) 851–858.

[20] R. Kessels, P. Goos, M. Vandebroek, A comparison of criteria to design efficient choice experiments, J. Market. Res. 43 (3) (2006) 409–419.

[21] F. Carlsson, P. Martinsson, Design techniques for stated preference methods in health economics, Health Econ. 12 (4) (2003) 281–294.

[22] Z. Sandor, M. Wedel, Profile construction in experimental choice designs for mixed logit models, Market. Sci. 21 (4) (2002) 455–475.

[23] N. Arora, J. Huber, Improving parameter estimates and model prediction by aggregate customization in choice experiments, J. Consum. Res. 28 (2) (2001) 273–283.

[24] A. Burton, D.G. Altman, P. Royston, R.L. Holder, The design of simulation studies in medical statistics, Stat. Med. 25 (24) (2006) 4279–4292.

[25] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159–174.

[26] D. Hensher, Reducing sign violation for vtts distributions through endogenous recognition of an Individual's attribute processing strategy, Int. J. Transp. Econ./ Rivista Int. di Econ. dei Trasporti 34 (3) (2007) 333–349.

[27] R.P. Merges, Uncertainty and the standard of patentability, High Technol. Law J. 7 (1) (1992) 1–70.

[28] L.W. Sumney, R.M. Burger, Revitalizing the U.S. semiconductor industry, Issues Sci. Technol. 3 (4) (1987) 32–41.

[29] Q. Liu, Y. Tang, Construction of heterogeneous conjoint choice designs: a new approach, Market. Sci. 34 (3) (2015) 346–366.

[30] B. Vermeulen, P. Goos, M. Vandebroek, Obtaining more information from conjoint experiments by best–worst choices, Comput. Stat. Data Anal. 54 (6) (2010) 1426–1433.

[31] J. Yu, P. Goos, M. Vandebroek, Efficient conjoint choice designs in the presence of respondent heterogeneity, Market. Sci. 28 (1) (2009) 122–135.

[32] U. Graßhoff, H. Großmann, H. Holling, R. Schwabe, Optimal design for discrete choice experiments, J. Stat. Plann. Inference 143 (1) (2013) 167–175.

[33] M. Crabbe, M. Vandebroek, Improving the efficiency of individualized designs for the mixed logit choice model by including covariates, Comput. Stat. Data Anal. 56 (6) (2012) 2059–2072.

[34] M.C.J. Bliemer, J.M. Rose, Construction of experimental designs for mixed logit models allowing for correlation across choice observations, Transp. Res. Part B Methodol. 44 (6) (2010) 720–734.

[35] B. Vermeulen, P. Goos, M. Vandebroek, Rank-order choice-based conjoint experiments: efficiency and design, J. Stat. Plann. Inference 141 (8) (2011) 2519–2531.

[36] B. Vermeulen, P. Goos, M. Vandebroek, Models and optimal designs for conjoint choice experiments including a no-choice option, Int. J. Res. Market. 25 (2) (2008) 94–103.

[37] B. Donkers, P.H. Franses, P.C. Verhoef, Selective sampling for binary choice models, J. Market. Res. 40 (4) (2003) 492–497.

[38] B.J. Kanninen, Optimal design for multinomial choice experiments, J. Market. Res. 39 (2) (2002) 214–227.

[39] J.F. Bridges, A.B. Hauber, D. Marshall, A. Lloyd, L.A. Prosser, D.A. Regier, F.R. Johnson, J. Mauskopf, Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force, Value Health 14 (4) (2011) 403–413.

[40] F.R. Johnson, E. Lancsar, D. Marshall, V. Kilambi, A. Mühlbacher, D.A. Regier, B.W. Bresnahan, B. Kanninen, J.F. Bridges, Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force, Value Health 16 (1) (2013) 3–13.

[41] V. Faustin, A.A. Adégbidi, S.T. Garnett, D.O. Koudandé, V. Agbo, K.K. Zander, Peace, health or fortune?: Preferences for chicken traits in rural Benin, Ecol. Econ. 69 (9) (2010) 1848–1857.

[42] B.K. Orme, Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research, Research Publishers, 2010.