



A Reporting Checklist for Discrete Choice Experiments in Health: The DIRECT Checklist

Jemimah Ride^{1,2} · Ilias Goranitis² · Yan Meng³ · Christine LaBond⁴ · Emily Lancsar⁴

Accepted: 18 August 2024 / Published online: 3 September 2024
© The Author(s) 2024

Abstract

Background Reporting standards of discrete choice experiments (DCEs) in health have not kept pace with the growth of this method, with multiple reviews calling for better reporting to improve transparency, assessment of validity and translation. A key missing piece has been the absence of a reporting checklist that details minimum standards of what should be reported, as exists for many other methods used in health economics.

Methods This paper reports the development of a reporting checklist for DCEs in health, which involved a scoping review to identify potential items and a Delphi consensus study among 45 DCE experts internationally to select items and guide the wording and structure of the checklist. The Delphi study included a best–worst scaling study for prioritisation.

Conclusions The final checklist is presented along with guidance on how to apply it. This checklist can be used by authors to ensure that sufficient detail of a DCE’s methods are reported, providing reviewers and readers with the information they need to assess the quality of the study for themselves. Embedding this reporting checklist into standard practice for health DCEs offers an opportunity to improve consistency of reporting standards, thereby enabling transparency of review and facilitating comparison of studies and their translation into policy and practice.

1 Introduction

Discrete choice experiments (DCEs) are a survey-based method widely used to ascertain individual preferences for health, healthcare services, and technologies and goods and services that affect health. The rapid growth of DCEs in health research in recent decades has resulted in methodological advancements and associated guidance on undertaking high-quality DCEs [1–5]. However, standards for the reporting of DCEs have lagged behind, with multiple recent systematic reviews highlighting the limited reporting of methodological detail and calling for better reporting standards to improve transparency, assessment

of validity, and translation of the results into policy and practice [6–8].

A key gap has been the lack of a consolidated checklist for the reporting of DCEs in the health literature. Reporting checklists, defined as ‘a checklist, flow diagram, or explicit text to guide authors in reporting a specific type of research, developed using explicit methodology’ [p. 1, 9], are commonly used for other methods in health economics, including economic evaluation [10], use of expert judgement [11], valuation of multi-attribute utility instruments [12] and value of information studies [13]. The purpose of such checklists is to improve reporting and its consistency, allowing readers to assess all important aspects of the study, and facilitating comparison across studies, such as in a systematic review. A reporting checklist differs from guidance documents or quality checklists [e.g. 1, 3–5, 14–16] in that it is not primarily aiming to improve the quality of the conducted study by guiding the researcher on how to carry out the methods, but rather to ensure that readers have all the information they need to assess the quality of the study for themselves.

The only existing reporting checklists applicable to health DCEs relate to the reporting of qualitative methods used in attribute development for a DCE, and aspects

Extended author information available on the last page of the article

Key Points for Decision Makers

Multiple reviews have called for better reporting standards for discrete choice experiments (DCEs) in health, highlighting inconsistent and insufficient reporting of methodological detail.

This study describes the development of a reporting checklist for DCEs in health using a scoping review to identify potential items, a Delphi consensus study among DCE experts to inform item selection and the wording and structure of the checklist, and piloting by a group of inexperienced DCE researchers who found it to be useful and easy to use.

The checklist is presented along with a guidance statement, with the goal that the checklist will become standard practice among DCEs in health, thereby improving transparency and assessment of study quality and enabling comparison of studies and their translation into policy and practice.

specific to online implementation of stated preference valuation surveys, of which DCEs are one type [17, 18]. Other reporting checklists not developed for DCEs in health are sometimes used in lieu of reporting checklists when one is required for submission, for example, Von Elm, Altman [19], as used in Xie, Liu [20]. To address this gap and ultimately improve reporting standards, the aim of this project was to develop a reporting checklist for DCEs in the health literature. We did this using a scoping review to identify potential items and an online Delphi consensus study among DCE experts.

2 Methods

This study was informed by guidelines for development of reporting guidelines in health [9], adapted to an online Delphi consensus study to allow for international participation and accommodate constraints arising from the coronavirus disease 2019 (COVID-19) pandemic. It also incorporated a best–worst scaling (BWS) study in round 2 of the Delphi study rather than the usual approach of using Likert-type responses. This was to overcome limitations of the latter, which include the number of items an individual can meaningfully rank, response biases and the lack of a theoretical basis for interpretation of the differences in scores. The project was registered with the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network as a reporting checklist in development [9]. Figure 1 presents an overview of the study's methods.

2.1 Scoping Review

A scoping review of the literature was conducted to identify potential checklist items. The eligibility criteria, search concepts, and terms and included databases are provided in the Supplementary material. Search terms were informed by previous systematic reviews [21–24]. Two reviewers conducted title and abstract screening in Rayyan [25] followed by full-text screening according to the eligibility criteria listed in Supplementary Table 1. The included papers (listed in Supplementary Table 2) included guideline documents on how to conduct a DCE, checklists to assess the quality of DCEs, checklists to report other types of stated preference studies, systematic reviews of DCEs, and specialised DCE taskforce reports on methods. From these, potential items were then reviewed independently by two members of the study team to identify duplicates or near duplicates across different sources, and further refined by the whole study team to exclude items that were only applicable to a small subset of DCEs, or that focussed on quality assessment rather than reporting. Reporting of the scoping review is limited by necessity due to the combination of reporting

Fig. 1 Overview of methods



with the rest of the study, and is therefore not in full accord with the relevant reporting guidelines [26].

2.2 Online Delphi Consensus Study Round 1

A key element in the development of reporting guidelines is consensus. The second stage of the project entailed an online Delphi consensus study among health economists and other experts conducting DCEs in health, as the most applicable set of people who will use the reporting checklist and have the expertise to identify the most relevant items. These methods were informed by guidelines for the conduct of Delphi studies [27] and by a previous exemplar study which used an online Delphi consensus survey to identify items for a standardised instrument for health economic evaluation [28]. Reporting was informed by guidelines for reporting of consensus studies [29].

Potential Delphi study participants were identified by a search for papers that included the terms “discrete choice experiment” and “health*” in the past 5 years in Web of Science, with authors ranked by the number of papers in this set. The first tranche to be invited were the top 50 ranked authors from this list (excluding the study team and those for whom an email address could not be sourced). Email addresses were sought from contact information included in publications, and from websites of universities or other employers, using affiliations listed in publications. To this list we added authors of key methodological papers in DCEs, authors of the guideline papers used in the scoping review, authors who were highly ranked by number of papers using the same search terms in Web of Science with no date restriction (to identify those with longer track records in the DCE literature) and editors of journals publishing health DCEs. There was the potential to invite further participants from the ranked list if needed. The target sample size was 30–50 participants, informed by Belton and MacDonald [27], who state that 20 participants may be sufficient to identify consensus, but that more may be useful, and the exemplar online Delphi [28], which had a sample of 45. Potential participants were invited by email and non-responders were followed up with twice. Consent was implied by completion of the survey, in which they were asked to agree to complete two rounds of the Delphi study. No financial incentive was offered for participation and invitees were advised that their responses would be anonymous. (See Supplementary material for the project protocol and explanatory statement provided to potential participants.)

Taking the list of items identified in the scoping review, we developed the round 1 online survey in Qualtrics (see Supplementary material for the full survey), which asked participants to consider each item in terms of how important it is to include in the reporting checklist on a scale of 1

(not at all important) to 7 (extremely important). The survey included opportunities to comment on items, to suggest items that participants thought were missing, and to provide input on the wording and structure of the items. The items were presented within domains representing the stages or components of the DCE process. Participants were also asked to nominate their top ten priority items for inclusion. Participants were asked about their experience of conducting DCEs, including the number of DCEs, their role(s), areas of application in health, and geographic areas in which they were applied.

2.3 Analysis of Round 1 of Delphi Study

Round 1 responses were analysed in terms of proportion scoring above and below cut-offs according to two pre-defined criteria (informed by Thorn and Brookes [28]) for items to be retained for round 2: (1) scored 6 or 7 by 50% or more of participants and 1–2 by less than 15% of participants; (2) included in top ten priority items by 15% or more of participants. Items not meeting these deliberately inclusive criteria were examined further to see whether consensus was clear that items could be dropped, including qualitative responses. Overlap with retained items was considered a reason for dropping. New items were added if suggested by 10% or more of participants. Qualitative comments were summarised by domain as to whether they captured suggestions on additional items or domains, wording changes or need for clarification, disagreement with inclusion of items, endorsement of items, and other comments.

2.4 Delphi Study Round 2

Those who participated in round 1 were invited to participate in round 2 of the Delphi study. In the invitation email for round 2, participants were sent a summary of round 1 results (see Supplementary material, which also includes round 2 survey text). Our approach for round 2 diverged from the study that informed our approach in round 1 [28], as we did not use Likert scale responses for round 2. Rather, the retained items were presented in two groups. Group 1 were items that met both criteria (1) and (2) above, while group 2 were items that only met criterion (1). Participants were advised that the group 1 items were to be included unless their responses indicated that any of them could be excluded, and asked to indicate whether each item should be kept or dropped. If participants elected to drop any item in group 1, they were asked to explain why. After being presented with the list of items in group 2, they were asked whether any of these items were important enough to include to be worth making the checklist longer to include them.

Following this, a BWS study was used to obtain a full rank of the items in group 2 [30]. The BWS was a case 1

(or object case) study in which participants were asked to choose the most important and least important items in an array of subsets of the items (choice tasks). The BWS used a 16-item balanced incomplete block design with 6 items per task, identified using the *find.BIB* function in R with the properties checked using the *isGYD* function. The design was randomly divided into two blocks, with participants randomly assigned to one of the two blocks upon entry to the study, so that each person completed eight tasks. The order of items within each task and the order of the tasks within each block were randomised per participant.

Participants were also asked what they considered to be the feasible maximum length of the checklist, and were advised that top-ranked items from group 2 may be included in the checklist depending on their responses to group 1 and the maximum survey length. They were also offered the chance to suggest re-wording of items in groups 1 and 2. Lastly, they were asked to review their earlier answer about the feasible maximum length of the checklist, having carefully considered all the items.

2.5 Analysis of Round 2 of Delphi Study

Items from group 1 were retained if at least 50% of participants nominated to keep them. Responses to the question about maximum feasible checklist length were free text, and were grouped into categories that emerged from the data. Qualitative content analysis was used to examine the reasons given by participants for dropping items from group 1 and suggestions made to change wording.

Group 2 items were analysed in terms of the proportion of participants stating that the item should be added to the checklist, the proportion of times that an item was selected as best, the proportion of times that the item was selected as worst, the mean individual best–worst score, the aggregate best–worst score, and the coefficients from multinomial logit models estimating the probability of the items being chosen as best and the equivalent model for probability of being chosen as worst. The best–worst score is the number of times an item is chosen as worst subtracted from the number of times it is chosen as best, either for an individual across the choice tasks they completed, or aggregated across the whole sample [31]. The multinomial logit model for probability of best took as its base level the item least often selected as best, and the model for probability of worst took as its base level the item least often selected as worst. We also examined a scatterplot of the mean and standard deviation (SD) of the individual best–worst scores [32]. This indicates consensus as well as the level of the score, with items in the lower right area of the scatterplot having higher mean scores (i.e. higher priority) and lower SD (i.e. greater consensus). Analysis of the BWS data was conducted in Stata V17 [33].

2.6 Final Checklist Selection

The study team met to decide the final list of items for inclusion in the checklist and to update wording of items in response to input from Delphi participants. The team considered each of the results from analysis of the round 2 results and also considered whether items were relevant to all contexts for DCEs in health, or whether they only applied to a subgroup of DCEs, and whether any items could be merged for conciseness. The inclusion of items from group 2 was considered in light of participants' comments on desirable checklist length and the ranking from the BWS study.

2.7 Piloting

The checklist was piloted among members of a DCE community of practice to which two of the study team belong. Group members were invited by email, and were asked to use the checklist to review reporting of a DCE paper of their choosing. Members were asked to provide feedback on the clarity of the checklist items, how easy it was to find all items in their chosen paper, which (if any) items were missing from the paper, formatting suggestions, and for details of the paper they reviewed (which could be published or a work in progress). Their feedback was summarised by topic and suggestions for changes to the checklist were considered by the study team.

This project was approved by the Monash University Human Research Ethics Committee (project ID 38369).

3 Results

3.1 Identification of Potential Checklist Items

In the scoping review, 1943 papers were title and abstract screened, resulting in 38 papers progressing to full-text screening, which identified 25 papers for inclusion (see Supplementary Table 2). From these, 313 potential items for the checklist were extracted. Two members of the team examined these items and removed duplicates, resulting in 162 unique items. Further refining of the list was conducted by the whole team on the following principles:

- Where items overlapped, they were combined if possible.
- Items were dropped if they were elements of good research practice and not specific to the conduct of a DCE (such as stating a clear research question).
- Items were also dropped if they related to quality of methods rather than reporting.

- The focus was on identifying items that apply to most DCEs in health, rather than being specific to a subtype of DCE.

This resulted in a final list of 48 items to be taken to the Delphi study (see Supplementary Table 3). The items were organised into seven domains that reflected the stages or components of a DCE life cycle, informed by the same documents used to extract potential checklist items. These items and domains are listed in the round 1 survey included in the Supplementary material.

3.2 Delphi Study Round 1 Results

Data collection for round 1 of the Delphi study was conducted between August and September 2023. The survey invitation was sent to 77 individuals, of whom 45 completed the survey, 7 declined to participate, and 25 did not complete the survey. There were seven individuals who we would have included in this first tranche of invitees, but we were unable to find a publicly available email address for them. One email address from the first tranche failed delivery and we were unable to find another email address for that person.

As presented in Table 1, the participating sample all had experience in reporting DCEs and most ($N = 41$, 91%) had been involved in six or more DCEs. The most common areas in which they had experience of applying DCEs were preferences for healthcare goods or services, preferences for goods or services that affect health and preferences for government policies, and their experience had been gained across all regions of the world.

As presented in Table 2, 41 of 48 items met criterion 1 (i.e. scored 6 or 7 by 50% or more of participants and 1–2 by less than 15% of participants). Of these, 23 also met criterion 2 (i.e. included in top 10 by 15% or more of participants). Seven items did not meet either criterion, and were therefore dropped. On the basis of the free-text responses, 17 items were re-worded, 5 items were combined into 2 items, and 1 new item was added ('Software used for analysis'). Other suggested additions related to a quality checklist rather than a reporting checklist or to general principles of good research practice and were therefore not added. After the amendments and re-wording, there were 23 items in group 1 (those meeting both criteria for retention) and 16 items in group 2 (those meeting only criterion 1, plus one new item).

Table 1 Experience in DCEs among the Delphi participants ($N = 45$)

		<i>N</i> (%)
How many DCEs involved in conducting	1–2	1 (2%)
	3–5	3 (7%)
	6+	41 (91%)
Which aspects of DCEs performed ^a	Interpretation	45 (100%)
	Reporting	45 (100%)
	Overall study design	43 (95%)
	Survey design	42 (93%)
	Experimental design	37 (82%)
	Data collection	37 (82%)
	Statistical analysis of choice responses	37 (82%)
	Qualitative research to select attributes and levels	35 (78%)
Areas of health and healthcare used DCEs ^a	Preferences for healthcare goods or services	44 (98%)
	Preferences for goods or services that affect health	32 (71%)
	Preferences for government policies	31 (69%)
	Quality of life	27 (60%)
	Health workforce preferences	18 (40%)
	Other	2 (4%)
Regions in which DCEs performed ^a	Europe	34 (75%)
	North America	24 (53%)
	Oceania	17 (38%)
	Asia	16 (35%)
	Africa	13 (29%)
	Latin America and the Caribbean	6 (13%)

^aParticipants could select multiple responses; therefore, figures can add up to more than 100%

Table 2 Scores from round 1 of the Delphi study

Domain	Item	Median score	% scoring 6–7	% scoring 1–2	% in top 10
Items relating to purpose and rationale	State the decision-maker perspective/s examined in the study	6	56%	0	27%
	Provide a rationale for using a DCE in the study	6	67%	2%	38%
Attributes and levels	Describe how attributes and levels were derived (e.g. systematic review, interviews, focus groups, expert input)	7	84%	0	49%
	Describe the process of iterative testing and refining of attributes and levels, including language	5	38%	0	4%
	Report attributes that were considered and excluded	5	33%	7%	4%
	Final list of attributes and levels	7	100%	0	91%
Experimental design	State the payment vehicle, if price included as an attribute	6	67%	2%	13%
	Describe how presentation of risk attribute/s was decided, if included	6	51%	4%	2%
	Indicate the number of alternatives per choice set	7	93%	0	27%
	Describe response options (e.g. forced choice, opt-out, status quo), with justification	7	99%	0	42%
	Report whether alternatives were labelled or unlabelled	7	91%	0	11%
	Report whether full or partial profile design	6	76%	2%	2%
	Describe the type of experimental design (e.g. full factorial, orthogonal, D-efficient, Bayesian efficient)	7	87%	0	42%
	Describe which effects are identified (main effects, higher order interactions, functional form)	7	82%	2%	16%
	Report the design properties, for example, D-efficiency, level balance, orthogonality	6	64%	4%	11%
	Report whether identification was checked (e.g. whether variance–covariance matrix block diagonal)	5	31%	9%	0
	Report whether design was blocked, and if so, how choice sets were allocated to blocks and whether properties of blocks were checked	6	71%	0	2%
	Describe the number of choice sets, number of blocks, number of choice sets per block	7	91%	0	36%
	Report whether some potential profiles were implausible and how this was addressed	6	56%	0	2%
	Report whether and how any a priori knowledge of signs and/or true parameters was used in the design	5	49%	2%	0
	Indicate how the design was obtained (software, catalogue, other)	6	78%	0	16%

Table 2 (continued)

Domain	Item	Median score	% scoring 6–7	% scoring 1–2	% in top 10	
Survey design and piloting	Report how respondents were allocated to blocks, if applicable	6	62%	2%	2%	
	Report other randomisation if used (e.g. choice task order, attribute order, alternative order, framing effects)	6	80%	0	4%	
	Provide the information, instructions, and questions seen in the survey (e.g. survey as an appendix)	7	80%	0	47%	
	Describe the medium used to communicate attribute/ level information (e.g. words, pictures, multimedia)	6	71%	0	2%	
	Describe visual implementation (colours, animation, text entry, drop-down menus, unique answering, scrolling design, etc.)	5	44%	4%	4%	
	Report pilot sample description and sample size	5	47%	7%	11%	
	Describe what was checked in piloting (e.g. understanding, respondent burden, timing, wording)	6	69%	0	11%	
	Report whether information from the pilot was used to update the experimental design (e.g. priors) and/or survey design	6	69%	0	7%	
	Sample and data collection	Report inclusion/exclusion criteria	7	89%	2%	40%
		Describe any use of quotas to ensure representativeness	6	71%	0	2%
Indicate the recruitment method (e.g. advertisement, invitation format, reminders)		6	71%	0	9%	
Describe how the target sample size was determined		6	58%	2%	9%	
Describe how data were collected (e.g. mail, personal interview, web survey)		7	93%	0	36%	
Report the response rate		6	62%	9%	20%	
Describe any incentives or remuneration for respondents		6	56%	0	2%	
If online – describe any methods used to avoid fraudulent responses (e.g. bots)		6	67%	0	0	
Report the final sample size		7	96%	0	49%	
Describe respondent characteristics and representativeness of target population		7	84%	2%	40%	
Econometric analysis	Indicate coding of data (effects/dummy/continuous)	7	87%	0	16%	
	Describe handling of missing data in choice tasks and/or other variables	7	78%	0	7%	
	Report whether any were responses were removed and why	7	93%	0	27%	
	Provide the rationale for model choice (e.g. conditional logit, mixed logit, GMNL, latent class, etc.) and assumptions (e.g. error variance)	7	87%	0	27%	
	Report model specification	7	89%	0	42%	

Table 2 (continued)

Domain	Item	Median score	% scoring 6–7	% scoring 1–2	% in top 10
Reporting of results	Report the model performance, goodness of fit	6	71%	0	20%
	Describe methods used for analysis of model results (e.g. calculation of marginal willingness to pay, attribute relative importance, welfare gain)	7	91%	0	51%
	Report the output/s of interest compared across a range of model specifications	5	44%	7%	13%
	Report measures of precision for the output/s of interest (e.g. confidence intervals) and how these were derived	7	76%	0	22%

Bolded cells indicate that the item met criteria for inclusion in round 2. Criterion 1: scored 6–7 by 50% or more of participants and 1–2 by less than 15% of participants; criterion 2: included in top ten priority items by 15% or more of participants. There was no criterion relating to the median score

3.3 Delphi Study Round 2 Results

Data collection for round 2 of the Delphi study was conducted between November 2023 and January 2024. All 45 participants from round 1 were invited to participate in round 2, of whom 38 (84%) completed responses. Characteristics of those retained for round 2 were similar to those who completed round 1 (see Supplementary Table 4). Regarding the feasible maximum length of the checklist, 55% ($N = 21$) indicated that more than 23 items would be acceptable (up to all 39 items), 24% ($N = 9$) stated that 23 items was about the right length, 8% ($N = 3$) wanted the length reduced to 20 items, and 13% ($N = 5$) provided no specific answer. When asked to reconsider this at the end of the survey, two revised their responses so that 60% ($N = 23$) accepted an increase in the checklist length from 23 items.

The maximum number of participants suggesting that an item from group 1 should be dropped was six, meaning none of these items were eligible to be excluded (on the basis of the predefined rule that an item would be retained if at least 50% of participants nominated to keep it), as presented in Table 3. The analysis of group 2 items was used for prioritisation, focussing on the top-ranking items according to the various criteria and the degree of consensus among the participants regarding each item. As presented in Table 4, reporting randomisation and describing what was checked in piloting were the most consistently high priority items across the different criteria, followed by reporting whether information from the pilot was used to update the DCE and describing how the sample size was determined. Figure 2 plots the mean and standard deviation of the individual best–worst scores. This shows that, of these four highly ranked items, the first three are clustered in the bottom right-hand corner of the chart, indicating that they have the highest mean scores with the greatest degree of consistency among the

participants. While the sample size and payment vehicle items have similar magnitude of mean individual best–worst scores, they have more variability in responses (as indicated by higher standard deviations), indicating less consensus. Notably, the item that was added to the round 2 survey on the basis of suggestions from round 1 (regarding software) had the lowest mean score and highest standard deviation, however, it was one of the most commonly suggested to add from group 2 (see column 1 of Table 4).

3.4 Piloting Results

The checklist was piloted by 19 members of a DCE community of practice, mostly researchers with limited experience of DCEs. The majority of feedback indicated that the checklist was easy to follow and helpful both for those reporting their own study and those reviewing a published paper, and that the examples provided were particularly helpful. In response to their feedback, the formatting of the checklist was amended to include item numbering and space to report the page number/s where an item may be found in the paper, and minor edits were made to wording. Their comments were also used to formulate the guidance on implementing the checklist (see Sect. 4.2).

3.5 Finalisation of the Checklist

The study team finalised the checklist on the basis of the quantitative analysis of the Delphi results, qualitative content analysis of the free text comments provided by participants and feedback from piloting. All 23 items from group 1 were kept, and the top 3 items from group 2 added (which all came under the survey design domain). In addition, the next two highest priority items from group 2 were incorporated into existing items from group 1 – the item regarding

Table 3 Delphi round 2 responses regarding dropping group 1 items

Item	<i>N</i> (%)
Describe the choice context and relevant decision-maker	1 (3%)
Provide a rationale for using a DCE to answer the research question	6 (16%)
Describe how attributes and levels were derived (e.g. systematic review, interviews, focus groups, expert input)	0
Provide the final list of attributes and levels	0
Report the number of alternatives per choice set	1 (3%)
Describe response options (e.g. forced choice, opt-out, status quo)	0
Describe the type of experimental design (e.g. full factorial, orthogonal, D-efficient, Bayesian efficient), whether full or partial profile design, and applicable design properties (e.g. D-efficiency, level balance, orthogonality)	1 (3%)
Describe which effects are identified (main effects, higher order interactions, functional form)	6 (16%)
Describe the number of choice sets, number of blocks, number of choice sets per block	0
Indicate how the experimental design was obtained (software, catalogue, other)	6 (16%)
Provide the information, instructions, a sample choice set and non-DCE questions seen in the survey (e.g. survey as an appendix)	2 (5%)
Report inclusion/exclusion criteria	0
Describe how data were collected (e.g. mail, personal interview, web survey)	0
Report the response rate, if known	3 (8%)
Report the final sample size	0
Describe respondent characteristics and representativeness of target population, if known	0
Indicate coding of data (effects/dummy/continuous) including definitions and cutoffs	3 (8%)
Report whether any were responses removed and why	1 (3%)
Provide the rationale for model choice (e.g. conditional logit, mixed logit, GMNL, latent class) and assumptions (e.g. error variance)	1 (3%)
Report model specification	1 (3%)
Report the model performance, goodness of fit if comparing models	4 (10%)
Describe methods used for analysis of model results (e.g. calculation of marginal rate of substitution, attribute relative importance, welfare gain)	0
Report measures of precision for the output/s of interest (e.g. confidence intervals) and how these were derived	3 (8%)

unlabelled/labelled alternatives was combined with the item on number of alternatives per choice set, and the item on how the sample size was determined was combined with the item on reporting the final sample size. This resulted in a total of 26 items in the final checklist, and 9 items were reworded because pilot participants indicated that these were unclear or ambiguous.

4 The DIRECT Checklist

The final checklist, the DIcrete choice experiment REporting ChecklisT (DIRECT), is presented in Table 5.

4.1 Definition of Terms Used in the Checklist

The terminology of several concepts in DCEs varies across the literature, thus for clarity their use in this checklist is defined here. Individuals completing the DCE are termed respondents. In the DCE, respondents are asked to choose between alternatives, each of which describes one instance of the good or service or health state in question. These alternatives are offered in choice sets (also termed choice scenarios

or choice tasks), with each choice set usually comprising two or more alternatives. The alternatives are described in terms of their attributes, which are the features or aspects of the alternatives for which respondents are hypothesised to have preferences. The attributes each have levels, which are values or categories that the attribute can take. Attributes may be termed dimensions, particularly when the DCE is used to value health states. Alternatives may be labelled, where the title of each alternative communicates information (such as a brand or type of intervention), or unlabelled, where the title is generic (such as ‘alternative A’ or ‘option 1’). Where the experimental design includes more choice sets than a single respondent can reasonably complete, the design is usually divided into blocks, with each respondent asked to complete only a single block of choice sets. Some DCEs include an opt-out alternative, in which the respondent may choose to take up none of the offered alternatives, or a status quo alternative, in which the respondent chooses the current situation rather than making a change by choosing one of the alternatives. Where respondents are not given an opt-out or status quo alternative, but have to choose one of the alternatives on offer, this is termed a forced choice DCE.

Table 4 Delphi round 2 analysis for group 2 items

Item	Add	Best	Worst	Individual BW score	Aggregate BW score	MNL—best	MNL—worst
	<i>N</i>	<i>N</i>	<i>N</i>	Mean	Mean	β	β
State the payment vehicle (e.g. out of pocket cost, additional tax), if price included as an attribute	10	30	15	0.39	15	1.76***	1.18*
Describe how presentation of any risk or probability attribute/s was decided	9	14	13	0.03	1	0.91	1.01
Report whether alternatives were labelled or unlabelled	14	25	15	0.26	10	1.55**	1.18*
Report whether design was blocked, and if so, how choice sets were allocated to blocks, whether properties of blocks were checked and how respondents were allocated to blocks	13	16	32	-0.42	-16	1.05*	2.06***
Report whether some potential profiles were implausible and how this was addressed	12	11	18	-0.18	-7	0.64	1.39**
Report any randomisation (e.g. choice task order, attribute order, alternative order, framing effects)	23	26	5	0.55	21	1.60**	[Base]
Describe the medium used to communicate attribute/ level information (e.g. words, pictures, multimedia)	11	21	15	0.16	6	1.36**	1.16*
Describe what was checked in piloting (e.g. understanding, respondent burden, timing, wording)	17	25	10	0.39	15	1.56**	0.75
Report whether information from the pilot was used to update the experimental design (e.g. priors) and/or survey design	11	26	9	0.45	17	1.59**	0.65
Describe any use of sample quotas	8	6	29	-0.60	-23	[Base]	1.89***
Indicate the recruitment method (e.g. advertisement, invitation format, reminders)	10	16	15	0.03	1	1.04*	1.15*
Describe how the sample size was determined	15	25	7	0.47	18	1.53**	0.36
Describe any incentives or remuneration for respondents	11	8	29	-0.55	-21	0.31	1.90***
If online – describe any methods used to avoid fraudulent responses (e.g. bots)	9	12	31	-0.5	-19	0.72	2.01***
Describe handling of missing data in choice tasks and/or other variables	12	26	16	0.26	10	1.59**	1.24*
Report the software used for analysis	20	17	45	-0.74	-28	1.09*	2.49***

Bold and italicized items are the top five most favoured in each column (in the case of ties both are highlighted)

Add number of participants stating the item should be included in the checklist, best number of times the item was rated as best (out of 288), worst number of times the item was rated as worst (out of 288), BW score best–worst score, MNL best and worst = coefficient from MNL model for choice of best and worst

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



Fig. 2 Scatterplot of mean and SD of individual BW scores

There may be different framing of the hypothetical choice scenario that respondents are asked to imagine.

4.2 Guidance on Implementing the Checklist

Here we provide guidance for using the checklist. Further detail on each item, including examples of how each has been met in published papers, is presented in the Supplementary material. The checklist is structured according to stages or components of conducting a DCE, covering purpose and rationale, attributes and levels, experimental design, survey design, sample and data collection, econometric analysis, and reporting of results.

The first items ask authors to provide readers with the purpose and rationale for the DCE, which set the scene for

Table 5 Checklist for reporting discrete choice experiments in health

Section Item		Page and paragraph
Purpose and rationale		
1	Describe the real-world context and decision-maker that the hypothetical choice context seeks to replicate or inform	
2	Provide a rationale for using a DCE to answer the research question	
Attributes and levels^a		
3	Describe how attributes and levels were derived (e.g. literature review, interviews, focus groups, expert input)	
4	Provide the final list of attributes and levels	
Experimental design		
5	Report the number of alternatives per choice set and whether they were labelled or unlabelled	
6	Describe response options (e.g. forced choice, opt-out, status quo)	
7	Describe the type of experimental design (e.g. orthogonal, D-efficient, Bayesian efficient, partial profile)	
8	Describe which effects are identified in the design (e.g. main effects, higher order interactions, functional form)	
9	Describe the number of choice sets, blocks and choice sets per block	
10	Indicate how the experimental design was obtained (software, catalogue, other)	
Survey design		
11	Provide a sample choice set and the instructions and background information given to respondents (e.g. providing the survey as an appendix)	
12	Report any randomisation (e.g. choice set order, attribute order, alternative order, framing effects)	
13	Describe what was checked in piloting (e.g. understanding, respondent burden, timing, wording)	
14	Report whether information from the pilot was used to update the experimental design (e.g. priors, functional form of attributes) or survey design	
Sample and data collection		
15	Report respondent inclusion/exclusion criteria	
16	Describe how data were collected (e.g. mail, personal interview, web survey)	
17	Report the response rate or cooperation rate, if possible	
18	Report the final sample size and how the sample size was determined	
19	Describe respondent characteristics and representativeness of target population, if known	
Econometric analysis		
20	Indicate coding of data (e.g. effects, dummy, continuous) including definitions	
21	Report whether any respondents were removed and why (e.g. suspected fraudulent responses, rationality tests)	
22	Provide the rationale for model choice (e.g. conditional logit, mixed logit, latent class) and assumptions (e.g. error variance)	
23	Report model specification	
Reporting of results		
24	Report the model performance, goodness of fit (if comparing models)	
25	Describe methods used for analysis of model results (e.g. calculation of marginal rate of substitution, attribute relative importance, welfare gain)	
26	Report measures of precision for the output(s) of interest (e.g. confidence intervals) and how these were derived	

^aFor reporting of formative qualitative analysis, see <https://www.equator-network.org/reporting-guidelines/using-qualitative-methods-for-attribute-development-for-discrete-choice-experiments-issues-and-recommendations/>

interpretation of the whole study. Item 1 asks for a description of the real-world choice at the heart of the research question, so that the reader can identify how well the DCE replicates this in the hypothetical context. In combination with transparency regarding the selection and characteristics

of the respondents (items 15 and 19), it also allows the reader to judge how well the sample represents the target population. However, Delphi participants pointed out that characteristics relevant to the research question, such as attitudes or experiences, may not be known for the target

population, limiting the analyst's capacity to assess relevant aspects of representativeness. Item 2 requires an explanation of why a DCE was a suitable approach to answer the research question, showing how the evidence produced by a DCE can be useful to decision-makers [5]. This may include why quantifying preferences is useful and/or why a DCE is preferred over other methods.

While we aimed to restrict the checklist to items that apply to most DCEs, it is not expected that all components of every item would be relevant to all studies. For example, not all DCEs will include all of the possible steps in developing attributes and levels (item 3). Sufficient detail should be provided on the approach taken to attribute development and sources of data to inform the selection of levels, so that the reader can assess their appropriateness. Where the development of attributes involved a systematic review or in-depth qualitative research, this may be published in a separate paper from the main DCE. A footnote to the checklist points the reader to existing reporting guidelines on qualitative methods used to develop attributes [17], which this checklist does not replicate or replace.

It may not be possible to include all checklist items in the main text of a paper, given word limits. It is sufficient to include items in Supplementary material, however, Delphi participants preferred the list of attributes and levels (item 4) to be included in the main paper wherever possible, as it is of fundamental importance for understanding the DCE.

Several items may be achieved by the inclusion of an exemplar choice set as a figure in the paper (item 11), showing how many alternatives each choice set contains and their titles (item 5), the response options (item 6) and wording of attributes (item 4). Inclusion of the survey text in an appendix allows the interested reader to see the detail of how the information provided to respondents was framed and to make assessments of whether the background information was sufficient or likely to introduce any unintended biases.

The checklist is not prescriptive in terms of the inclusion of specific effects in the design (item 8), nor that the design has to exactly match what is estimated in the model. Rather, this reporting will enable the reader to make a judgement about the validity of the model estimated from the choice data, in combination with the items on type of design (item 7), model specification (item 23) and sample size (item 18). For example, if the estimated model includes two-way attribute interactions and/or non-linear functional forms of continuous attributes on a modest sample size, the reviewer may assess this differently if they are informed that the design included only linear main effects or if all those effects were identified in the design.

For clarity, the checklist asks that some optional methods be reported even when this is to report that they were

not done. It is not necessary that all DCEs involve randomisation (item 12) in the presentation of the survey to respondents, and there may good reasons not to randomise, but knowing whether randomisation has been used allows the reader to assess potential for issues that could be impacted by randomisation. This can include randomisation to different versions of the survey or blocks, ordering of choice sets or alternatives, and attribute order within choice sets. Reporting should specify whether randomisation was within or between individual respondents. Items 13 and 14 relate to piloting (what was checked and how information from the pilot was used), which may not be included in all studies but were high priority to include from the BWS study.

It may be difficult to report a response rate for DCEs (item 17), depending on the recruitment method. When recruitment involves advertising on a website or social media, or via a third-party survey company, it may not be possible to know how many people saw the invitation in total. However, it may still be informative to provide what is known about the number of completed responses compared with the number of people who clicked a link or opened the survey (completion rate), or the proportion of those who were invited to participate who made a partial or complete response (cooperation rate) and/or those who declined to consent or dropped out part way.

There are different approaches to determining the target sample size for DCEs [1, 34, 35]. Item 18 does not require the reporting of formal sample size calculations but asks for information on how the sample size was determined to allow the reader to interpret its appropriateness for the study. It can be useful to describe how the sample size relates to the size of the target population (for example, if the entire target population is a small group of national decision-makers, or all taxpayers in the country).

A growing issue of concern in the DCE research community is that of fraudulent or invalid responses. After data collection, data from some respondents may be removed from the dataset due to red flags suggesting that responses do not represent a real individual or that the respondent was not sufficiently engaged or misunderstood the DCE. A range of strategies may be used to investigate these possibilities, such as repeating demographic questions in different ways, hidden questions that are not visible to human respondents, dominance tests, analysis for straight lining and how long it takes to complete the survey [36, 37]. Item 21 does not suggest that all DCEs require these tests, but where they have been conducted and used to drop suspicious responses, this should be reported. This is an emerging field—the reporting checklist for online stated preference studies by Menegaki and Olsen [18] from 2016 only asks how respondents were

prevented from responding multiple times and whether a minimum completion time was imposed. Transparency of reporting is therefore vital to allow readers to see how these risks were handled.

Depending on the audience, it is not always necessary or appropriate to include the model equation in the paper, provided that the model specification (item 23) is clear from what is reported. When multiple model specifications are presented, authors are asked to report measures such as log-likelihood, pseudo *R*-squared, likelihood ratio tests, Akaike information criteria (AIC) and/or Bayesian information criteria (BIC) to assist the reader in interpreting the choice of main model (item 24). However, these measures may not be the primary driver of model choice [1], and authors are also asked to provide their reasons for choosing a particular modelling approach with clarity on its assumptions (item 22). Apart from reporting model coefficients, DCE papers often report further analysis such as those outlined in item 25 [1, 38]. The methods used to generate these outputs should be reported, along with estimates of uncertainty and how these were obtained (item 26) [39].

5 Discussion

This study produced the DIRECT consolidated checklist for reporting of health DCEs, informed by a scoping review, a Delphi consensus study among experienced DCE researchers (with a BWS study to prioritise items for inclusion) and feedback from less experienced DCE researchers on the usability of the checklist. We propose that this checklist become a standard part of reporting for health DCEs to address the need for improved and standardised reporting of DCEs by providing a concise list of items to report that will allow for scrutiny of the methods used. The checklist focusses on issues specific to DCEs rather than common principles of good research practice, and on items that were considered relevant to most DCEs in health. The intent is not that this checklist should inform study authors on how to perform a DCE; for that purpose, we would refer the reader to existing guidelines. Nor has this checklist been developed to produce a score of reporting quality. Along with the authors of the CHEERS checklist [10], we would caution that such use could be misleading. For example, it is not clear that all items in the checklist should have the same weight if a score of reporting quality was being produced. We developed this checklist specifically for use with DCEs, so some items are not relevant to other types of stated preference studies.

The development of this checklist drew from the literature review and the input of highly experienced and inexperienced DCE researchers, providing a well-rounded set of perspectives. While we built consensus through the Delphi

study, we note that participants' responses did indicate differing perspectives on how extensive and detailed the checklist should be. For example, in round 1, participant comments ranged from 'I can't imagine a paper being accepted without all this!' to 'I think a lot (most) of this should be in an appendix' to 'Combine elements to more global questions'. In our approach to checklist development, the final selection of the checklist rested with the study team, rather than being a consensus decision with the Delphi study participants. However, feedback was proactively sought iteratively through the different phases of the development of the checklist, and final decisions about items and wording were always in alignment with the broad consensus among the Delphi participants and the objective criteria specified a priori. The number of participants in the Delphi was similar to the sample size of previous similar studies, such as Thorn and Brookes [28], but relatively small compared with other BWS studies [40].

We did not specify where each item should be reported in the manuscript, as some reporting checklists have done [e.g. 10]. This was to allow flexibility, since DCEs in health are published in journals with differing norms, such as word counts or the suitability of including details such as econometric items in clinically oriented journals. In both round 2 of the Delphi and in piloting, we asked for feedback on the format of the checklist, and while there were no suggestions to include guidance on placement of items in the manuscript, participants did note that many items could be reported in an appendix, or via the provision of the survey text or an exemplar choice set, which we have noted in the guidance.

Delphi studies aiming to gain consensus on prioritisation have used various methods to measure and establish consensus, most often using Likert scales and a defined percentage of agreement [41], as we used in round 1 of this study along with nomination of the top ten items. For round 2, we used a case 1 BWS to obtain a ranking of the group 2 items for inclusion in the checklist, following guidelines for such studies [30]. The advantages of BWS include its foundation in random utility theory, that it accommodates more items for ranking than would be feasible using other methods such as full ranking or pairwise comparison and that it directly elicits priorities [30, 31, 40]. We found that combining these quantitative methods with qualitative analysis of free-text responses to be valuable and efficient in refining the checklist down from the original 48 potential items to the final 26 items.

5.1 Future research

The usefulness of this checklist will be evident from whether it becomes part of standard practice and whether the reporting quality of the health DCE literature improves

over coming years, with future systematic reviews of this literature able to compare with reporting quality in those conducted to date [e.g. 6–8, 22]. We invite feedback on the checklist that could be used to improve future iterations. Next steps for this checklist will be to disseminate the checklist among the health preference research community and to develop and test an accompanying evaluation tool that is designed to generate a reporting score [42].

6 Conclusions

This paper reports the development of the DIRECT checklist for reporting of DCEs in health. We propose that use of this reporting checklist become standard practice, with the aim of improving reporting standards and thereby facilitating better transparency, assessment of validity and translation into policy and practice.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-024-01431-6>.

Acknowledgements The authors express their heartfelt gratitude to all of the Delphi study participants who contributed to the development of this checklist, to those who provided comments when the checklist in development was presented at the Australian Health Economics Society conference 2023, and members of the DCE community of practice who piloted the checklist. We also thank one of the Delphi participants for the suggestion to include best–worst scaling as a method of prioritisation.

Declarations

Funding Open Access funding enabled and organized by CAUL and its Member Institution. Jemimah Ride's time on this project was partly funded by the Elizabeth and Vernon Puzey Early Career Research Fellowship at the University of Melbourne.

Conflicts of Interest The authors have no financial or proprietary interests in any material discussed in this article.

Availability of Data and Material Some data (such as de-identified and aggregated data, in line with ethics approval) may be available upon request to the authors.

Ethics Approval This project was approved by the Monash University Human Research Ethics Committee (project ID 38369).

Author Contributions All authors contributed to the study conception and design. Yan Meng performed the scoping review, with contribution of all authors to refine the list of items. Jemimah Ride conducted the Delphi study with input from all authors. All authors were involved in formulating the checklist and refining it at each stage of development. The original draft of the manuscript was prepared by Jemimah Ride and reviewed and edited by all authors. All authors reviewed and approved the final draft.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction

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

References

1. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *Pharmacoeconomics*. 2008;26(8):661–77.
2. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swannutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ*. 2012;21(6):730–41.
3. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health*. 2016;19(4):300–15.
4. 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.
5. Bridges JF, de Bekker-Grob EW, Hauber B, Heidenreich S, Jansen E, Bast A, et al. A roadmap for increasing the usefulness and impact of patient-preference studies in decision making in health: a good practices report of an ISPOR task force. *Value Health*. 2023;26(2):153–62.
6. Karim S, Craig BM, Vass C, Groothuis-Oudshoorn CGM. Current practices for accounting for preference heterogeneity in health-related discrete choice experiments: a systematic review. *Pharmacoeconomics*. 2022;40(10):943–56.
7. Merlo G, van Driel M, Hall L. Systematic review and validity assessment of methods used in discrete choice experiments of primary healthcare professionals. *Health Econ Rev*. 2020;10(1):39.
8. Vass C, Rigby D, Payne K. The role of qualitative research methods in discrete choice experiments: a systematic review and survey of authors. *Med Decis Making*. 2017;37(3):298–313.
9. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2): e1000217.
10. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 explanation and elaboration: a report of the ISPOR CHEERS II Good Practices Task Force. *Value in Health*. 2022;25(1):10–31.
11. Iglesias CP, Thompson A, Rogowski WH, Payne K. Reporting guidelines for the use of expert judgement in model-based economic evaluations. *Pharmacoeconomics*. 2016;34(11):1161–72.
12. Xie F, Pickard AS, Krabbe PFM, Revicki D, Viney R, Devlin N, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–77.
13. Kunst N, Siu A, Drummond M, Grimm SE, Grutters J, Husereau D, et al. Consolidated Health Economic Evaluation Reporting

- Standards-Value of Information (CHEERS-VOI): explanation and elaboration. *Value in Health*. 2023;26(10):1461–73.
14. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health*. 2011;14(4):403–13.
 15. Joy SM, Little E, Maruthur NM, Purnell TS, Bridges JF. Patient preferences for the treatment of type 2 diabetes: a scoping review. *Pharmacoeconomics*. 2013;31:877–92.
 16. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC Health Serv Res*. 2014;14(1):1–14.
 17. Hollin IL, Craig BM, Coast J, Beusterien K, Vass C, DiSantostefano R, et al. Reporting formative qualitative research to support the development of quantitative preference study protocols and corresponding survey instruments: guidelines for authors and reviewers. *The Patient Patient-Cent Outcomes Res*. 2020;13(1):121–36.
 18. Menegaki AN, Olsen SB, Tsagarakis KP. Towards a common standard—a reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *J Choice Model*. 2016;18:18–50.
 19. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 2007;370(9596):1453–7.
 20. Xie Z, Liu H, Or C. A discrete choice experiment to examine the factors influencing consumers' willingness to purchase health apps. *Mhealth*. 2023;9:21.
 21. 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.
 22. 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.
 23. Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB, et al. Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. 2012.
 24. Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol*. 2011;64(7):718–42.
 25. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
 26. Tricco A, Straus S, Moher D. Preferred reporting items for systematic reviews and meta-analysis: extension for scoping reviews (PRISMA-ScR). *Ann Intern Med*. 2018;169(7):467–73.
 27. Belton I, MacDonald A, Wright G, Hamlin I. Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technol Forecast Soc Chang*. 2019;147:72–82.
 28. Thorn JC, Brookes ST, Ridyard C, Riley R, Hughes DA, Wordsworth S, et al. Core items for a standardized resource use measure: Expert Delphi Consensus Survey. *Value in Health*. 2018;21(6):640–9.
 29. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, et al. ACCORD (ACcurate CONsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med*. 2024;21(1):e1004326.
 30. Strasser A. Design and evaluation of ranking-type Delphi studies using best-worst-scaling. *Technol Anal Strategic Manag*. 2019;31(4):492–501.
 31. Louviere J, Lings I, Islam T, Gudergan S, Flynn T. An introduction to the application of (case 1) best–worst scaling in marketing research. *Int J Res Mark*. 2013;30(3):292–303.
 32. Aizaki H, Fogarty J. R packages and tutorial for case 1 best–worst scaling. *J Choice Model*. 2023;46:100394.
 33. StataCorp. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC; 2021.
 34. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in health-care: a practical guide. *The Patient Patient-Cent Outcomes Res*. 2015;8(5):373–84.
 35. Orme B. Sample size issues for conjoint analysis studies. Sequim: Sawtooth Software Technical Paper. 1998.
 36. 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.
 37. Pearce A, Harrison M, Watson V, Street DJ, Howard K, Bansback N, et al. Respondent understanding in discrete choice experiments: a scoping review. *The Patient Patient-Cent Outcomes Res*. 2021;14(1):17–53.
 38. Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. *Soc Sci Med*. 2007;64(8):1738–53.
 39. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ*. 2007;16(8):827–40.
 40. Hollin IL, Paskett J, Schuster ALR, Crossnohere NL, Bridges JFP. Best-worst scaling and the prioritization of objects in health: a systematic review. *Pharmacoeconomics*. 2022;40(9):883–99.
 41. Niederberger M, Spranger J. Delphi technique in health sciences: a map. *Front Public Health*. 2020;8.
 42. Logullo P, MacCarthy A, Kirtley S, Collins GS. Reporting guideline checklists are not quality evaluation forms: they are guidance for writing. *Health Sci Rep*. 2020;3(2)e165.

Authors and Affiliations

Jemimah Ride^{1,2}  · Ilias Goranitis²  · Yan Meng³ · Christine LaBond⁴  · Emily Lancsar⁴ 

✉ Jemimah Ride
jemimah.ride@monash.edu

¹ Monash University Health Economics Group, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

² Health Economics Unit, Melbourne School of Population and Global Health, University of Melbourne, Parkville, Australia

³ Analysis Group Ltd, London, UK

⁴ Department of Health Economics Wellbeing and Society, The Australian National University, Canberra, Australia