Article:

Attribute selection for a discrete choice experiment incorporating a best-worst scaling survey

Edward J.D. Webb[*,†,‡]   David Meads[*†]   Yvonne Lynch[§]

Simon Judge[§**]   Nicola Randall[§,**] Juliet Goldbart[§]   Stuart Meredith[§]   Liz

Moulam[§]   Stephane Hess[†,††] Janice Murray[§]

Acknowledgements

Funding

Ethical approval

Ethical approval was received for the study from an NHS Research Ethics Committee (REC reference 6/NW/0165) and informed consent was obtained from participants at the start of the survey.

Data availability

---

[*] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[†] Choice Modelling Centre, University of Leeds, Leeds, UK
[‡] Corresponding author; e.j.d.webb@leeds.ac.uk
[§] Faculty of Health, Psychology and Social Care, Manchester Metropolitan University, Manchester, UK
[**] Barnsley Assistive Technology Team, Barnsley Hospital NHS Foundation Trust, Barnsley, UK
[††] Institute for Transport Studies, University of Leeds, Leeds, UK

Survey data is not publicly available as respondent consent was not obtained for this. However, it is available on request to the corresponding author or to Leeds Institute of Health Sciences if a formal data sharing agreement is entered into.

Abstract

Background: While literature exists on using qualitative methods to generate potential attributes for a discrete choice experiment (DCE), there is little on selecting which attributes to include.

Aim: To present a case study in which a best-worst scaling case 1 (BWS-1) survey was used to guide attribute selection for a DCE. The case study's context was the decision making of professionals around the choice of augmentative and alternative communication (AAC) systems for children with limited natural speech.

Methods: BWS-1 survey attributes were generated from literature reviews and focus groups. DCE attributes were selected from BWS-1 attributes. The selection criteria were: include mostly important attributes; create coherent descriptions of children and AAC systems; address the project's research aims; have an appropriate respondent burden. Attributes' importance was judged using BWS-1 relative importance scores.

Results: The BWS-1 survey included 19 child and 18 AAC device/system attributes and was administered to N=93 AAC professionals. Four child and five device/system attributes were selected for the DCE, administered to N=155 AAC professionals.

Conclusions: In this case study BWS-1 results were useful in DCE attribute selection. Four recommendations are made for future studies: define selection criteria for DCE attributes a priori; consider the impact participant's perspective will have on BWS-1 and DCE results; clearly define key terminology at the start of the study and refine it as the study progresses to reflect interim findings; BWS will be useful when there is little existing stated preference work on a topic and/or qualitative work is difficult.

Keywords: discrete choice experiment; best-worst scaling; attribute development; attribute selection; methodology

1. Introduction

Discrete choice experiments (DCEs) are an increasingly popular tool to investigate preferences in healthcare.[1-4] In a DCE respondents make choices between hypothetical objects (e.g. treatments) described by a set of attributes (e.g. efficacy, side effects). Each attribute can take a number of levels, and by varying the levels in each question, it is possible to analyse the trade-offs respondents make between them.

Developing a DCE requires constructing a set of attributes and levels, which are meaningful to respondents, capture the most salient aspects of decision-making and can address the research question. It is usually best practice to generate attributes using qualitative methods (e.g. focus groups) as well as drawing on previous literature and expert input.[5-11] However, this process usually produces more potential attributes than the 5-6 typically included in a DCE.

The process of selecting final attributes from a list of many candidates is relatively neglected in the literature. There is no consensus on the optimal way to develop DCEs,[12] but Janssen et al.[13] propose a general framework which has been used by several DCEs[14-18]. However, none of the framework's five steps[*] explicitly include selecting which attributes to include from a list of candidates. Bridges et al.[5] present a checklist for good practice in DCEs which includes whether attribute selection was done appropriately (item 2.2). However, they do not discuss what criteria should be used for selection, or give details about what methods are useful in practice.

We present a case study in which as part of a single pre-planned research project a best-worst scaling case 1[†] (BWS-1) survey[19] was conducted followed by a DCE,[20] with BWS-1 results used to inform DCE attribute selection. This study sought to demonstrate that BWS-1 results can provide a sound and quantitatively robust evidence base for researchers to draw on in DCE attribute selection. We present a case study, highlighting issues which arose, discussing advantages and disadvantages of the process, and finish with a list of recommendations for future studies.

BWS-1, and the related methods of BWS case 2 (BWS-2) and BWS case 3 (BWS-3) surveys are common in healthcare.[21] Participants are shown 5-6 attributes drawn from a longer list and select which they consider the best and worst (or alternative terms depending on context). The relative importance of each attribute can then be estimated on a ratio scale. BWS-1 contrasts with DCE, as participants choose a single attribute in isolation, rather than selecting a choice object comprised of multiple attributes. BWS-1 is complementary to DCE, as it can accommodate many more attributes (e.g. Hofstede et al.[22] include 53), though at the cost of not gathering as much information on any single attribute. BWS-2 and BWS-3 may be seen as substitutes for DCE, as all three methods examine a similar number of attributes, with different levels for the attributes.

A literature search was carried out (see appendix for details). Several studies compare BWS and DCE results, the majority of which focus on BWS-2,[23-29] although some studies either compare or combine BWS-1 and DCE

---

[*] The five steps are: evidence synthesis, expert consultation, stakeholder engagement, pre-test interviews and pilot testing.
[†] Also known as best-worst scaling object case.

results.[30-33] However, no study used BWS-1 results in DCE attribute selection, thus we believe this is the first study to do so.

DCE and BWS-1 results may have different properties, with one interpretation being that BWS elicits relative values rather than preferences.[34] However, data from each method can be analysed using a random utility framework,[35] and there is evidence that BWS-1 results are consistent with participants' preferences in a DCE.[30-33]

The case study was carried out in the field of augmentative and alternative communication (AAC) provision for children as part of a wider research project called Identifying appropriate symbol communication aids for children who are non-speaking: enhancing clinical decision making (I-ASC).[36] Children with speech difficulties rely on AAC to express themselves. Speech difficulties may be related to several conditions (cerebral palsy, intellectual/developmental delays, autism spectrum condition). AAC encompasses many aids and techniques which supplement spoken communication, including low-tech options such as picture/letter boards and high-tech devices which may synthesize speech. The needs and abilities of children vary hugely, and AAC systems/devices[*] must be carefully matched to children to take account of each individual's particular needs and circumstances.

No prior stated preference work existed in AAC to give guidance as to which potential attributes to select. In addition, the complex nature of the decision-making environment meant that we wished to present DCE choice tasks with participants choosing between which AAC system to provide for a hypothetical child vignette, constructed from a set of child attributes. We thus required attributes both for children and AAC systems, contributing to there being many potential attributes, with little evidence as to which to select. The above situation was the primary motivation to conduct a BWS-1 survey, then incorporate the results into DCE attribute selection.

The present article focuses on survey instrument development. For details of the individual studies, see Webb et al.[19] and Webb et al.[20]. For more information on AAC terminology, Beukelman and Mirenda[37] is an introductory text.

2. Methods

Figure 1 gives a timeline of project events.

Focus groups were conducted with 31 AAC clinical specialist stakeholders [38,39] and reviews of the AAC literature were carried out.[40] Potential BWS-1 attributes were identified from this material and condensed into an initial list of child and AAC device attributes by authors (EW, DM, YL, NR, SJ, JG, LM) with expertise in AAC, speech and language therapy, qualitative research and health economics. These lists were reduced by combining similar attributes (e.g. age and educational stage). Feedback was received from other authors and a project advisory group comprising AAC professionals, people who use AAC, and relatives and support personnel of people who use AAC. Attribute identification and refinement was done by unstructured discussion between authors until all were in agreement.

---

[*] Note that the terms "system" and "device" are not fully synonymous, and device was used in the BWS-1 survey and system was used in the DCE survey. Further details are given in section 3. In the current paper, the word system is used except for in the specific context of the BWS survey.

Two BWS-1 exercises, one for child and one for AAC device attributes, were administered online as a single questionnaire to AAC professionals who work with children. The questionnaire was piloted with N=5 AAC professionals and minor changes made to language and formatting. Figure 2 shows an example task. Results were analysed using mixed logit models, producing relative importance scores (RIS) for attributes.[19]

Following the BWS-1, a DCE was carried out. Respondents chose an AAC device/system for a vignette describing a hypothetical child. Child and AAC system attributes were selected from BWS-1 attributes. This was achieved by consensus discussions between all authors.

Attributes were selected using the following criteria:-

1) Most attributes should be anticipated to be highly important in participants' decision-making;
2) DCE results would address the research project's specific aims;
3) DCE attributes should form coherent and realistic descriptions for both children and AAC systems;
4) The number of attributes should be small enough that the cognitive burden would be reasonable, and that models could be estimated given the expected sample size.

Attributes' anticipated importance (criterion 1) was judged using attributes' RIS from the BWS-1. Overall, attributes were selected pragmatically: no explicit weighting was given to the criteria, but individual attributes contributed more to satisfying some criteria than others.

Although BWS-1 RIS was important in selecting attributes, there was no pre-set threshold above which an attribute would automatically be selected (e.g. if an attribute was in the top three it would be selected). The research project particularly aimed to investigate symbols used in AAC systems, thus attributes related to graphic symbols were prioritised (criterion 2). However, no a priori decisions were made to include any attributes in the DCE regardless of BWS-1 results. Whether DCE attributes satisfied criteria 3 and 4 was assessed firstly by members of the research team with AAC expertise, then verified in piloting.

The attributes' language was refined in the light of BWS-1 findings and to fit a different decision-making context. (For example, in a DCE levels are presented alongside attributes, unlike in BWS-1.)

Attribute selection and refinement was carried out by unstructured discussion between authors until consensus was achieved. A DCE was created and piloted with N=5 AAC professionals, with minor changes made to wording and formatting. Figure 3 shows an example choice task. The DCE was administered online to AAC professionals.[20]

3. Results

Initial lists of 31 child and 29 AAC device related attributes (Table A 1 and Table A 2) were condensed into lists of 19 child and 18 AAC device attributes (Table 1) used in the BWS-1. Completed responses from N=93 participants were analysed, producing meaningful results.[19]

Following the BWS-1, four child and five AAC system attributes were selected for the DCE, given in Table 3, as well as attributes' rank in terms of BWS-1 RIS. The wording of three out of four child attributes and three out of six AAC system attributes was edited between BWS-1 and DCE.

Predicted future needs and abilities had the sixth highest RIS. However, it was felt that for vignettes to give a realistic summary of a child, it was necessary to include an indication of the child's predicted future trajectory.[41,42] "Children's predicted future needs and abilities" was changed to "predicted future skills and abilities" to make levels easier to create. Levels needed to capture a child progressing or regressing in ability, and their needs changing due to different abilities. However, it was difficult to succinctly capture changing needs due to developing/regressing abilities, so the wording was changed to remove the reference to needs.

Graphic representation ranked 12th out of 18 in RIS, but was included as one of five AAC system attributes, due to the wider research project's focus on graphic representation (the type of symbol a system uses).

Some changes in wording were made due to the different formats of BWS-1 and DCE. The layout of the DCE (Figure 3) displayed AAC system attributes as short "headlines" followed by longer descriptions below. This was possible due to attribute labels remaining unchanged in each DCE question, so respondents did not have to read them every time. In the BWS-1, different attributes were shown in each question, and respondents had to read the whole list every time, making a similar visual layout to the DCE impractical. Thus the information contained in the DCE attribute descriptions had to be included in the attribute name for the BWS-1 survey. For example, "vocabulary or language packages" in the BWS-1 became "vocabulary sets" in the DCE.

The term "AAC device" used in the BWS-1 was changed to "AAC system". Although nomenclature is not completely rigid, participants perceived "AAC device" to imply high-tech powered electronic devices, whereas AAC system encompassed both powered high-tech devices as well as paper and person based AAC strategies, for example boards with letters/words/pictures on them. The change was possible as attributes specific to high-tech powered devices, e.g. battery life, were not included in the DCE.

DCE pilot testing revealed a need for a slightly different target population than for the BWS-1. Testing showed some AAC professionals, although they had clinical experience, had little input into decisions about which AAC system to choose for a child. (For example, occupational therapists who help children with adjustments to AAC systems selected by other members of a clinical team.) These individuals struggled to give meaningful answers drawing on their own experience. No similar problems were found in BWS-1 testing. A screening question was added to the survey which redirected individuals without the necessary experience for completing the DCE tasks to answer only demographic questions.[*]

The survey was completed by N=172 participants, of which 155 answered yes to the above question and answered DCE questions. Analysis produced meaningful results.[20]

4. Discussion

---

[*] The specific wording of the questions was: *"I confirm my work involves assessing children for AAC systems and I contribute to the decision making in relation to the language and vocabulary organisation within AAC systems (Yes/No)."*

The DCE attribute selection criteria detailed in section 2 were similar to those implicitly or explicitly used in developing most DCEs (see e.g. item 6 of the ISPOR Checklist for Conjoint Analysis in Health[5]). The difference here is that candidate attributes' anticipated importance was informed by BWS-1 results, and it is instructive to examine how this influenced the application of the criteria.

We did not simply select the top ranked BWS-1 attributes, as this would not have resulted in an appropriate DCE instrument. The attributes graphic representation and battery life illustrate how different selection criteria were more relevant for different attributes. There was no significant difference in the attributes' RIS, being ranked 12th and 13th respectively out of 18. Although neither was expected to be among the most important in DCE participants' decision-making (criterion 1), graphic representation was the attribute most suited to address the project's focus on symbols used in AAC systems. Thus it was selected for inclusion under criterion 2. Battery life was not useful to address the project's aims, nor was it necessary for a realistic description of an AAC system (criterion 3). Thus it was excluded from the DCE, reducing the number of attributes in line with criterion 4.

There was less reliance on the research team's prior beliefs when applying criterion 1 than would otherwise have been the case. The authors' prior belief was that battery life would be important, and there were even worries that it might dominate other attributes. However, the BWS-1 provided detailed quantitative information from many AAC professionals indicating battery life did not fulfil the inclusion criteria.

For some attributes, criterion 3 was most relevant. For example, predicted future skills and abilities, with the 6th highest RIS, was included as one of four child attributes, as it was necessary to create realistic child vignettes. As four child attributes was felt to be a maximum to not overburden participants (criterion 4), this necessitated not including one of the top four ranked child attributes. Support for AAC from communication partners, ranked 2nd, was omitted as it was felt to be the least necessary to form realistic child vignettes or to address the project's aims.

The above discussion highlights that selecting DCE attributes remains a pragmatic and subjective process after incorporating a BWS-1. However, BWS-1 results can give clarity to both researchers and the reader. As the BWS-1 results were also published, the reader can observe researchers' selection decisions, in line with recommendations[43] that studies report greater detail on instrument development.

It was assumed that the relative importance participants ascribed to BWS-1 attributes was informative about how DCE participants would respond. BWS-1 and DCE have different properties. It is argued that they elicit different sorts of information, with DCEs eliciting preferences and BWS values,[24,34] and there is evidence they measure different constructs.[23] However, a random utility framework is usually used to analyse both BWS-1 and DCE data,[35,44] and there is evidence that BWS-1 RIS corresponds to utility weights in a DCE.[30-33]

BWS-1 results gave additional information to update prior beliefs about which candidate attributes fulfilled the criteria for inclusion in the DCE. In this light, methodological differences seem less important, as the assumption that BWS RIS is informative about the relative importance of DCE attributes is both reasonable and supported by empirical evidence,[33] not least the current project's results. For example, graphical representation had the lowest RIS of any BWS-1 attribute included in the DCE. Although it had a significant effect on decision-making in the DCE, it also had the lowest relative importance.

BWS-1 is an alternative to existing methods which also have significant differences in their properties and the constructs they measure. Studies using nominal group technique often perform a ranking exercise at the end [6]. Others studies use Q-methodology, in which participants sort candidate attributes from most to least important on a board.[45,46] BWS-1 has advantages over these alternatives, for example it imposes a lower cognitive burden than ranking, [47] potentially giving more reliable results, as well as allowing statistical significance testing of the difference in importance of two attributes.

The conclusions that drawn from BWS-1 and DCE depend on their context, and on what perspective participants are asked to complete tasks from.[48,49] The current study examined a particular perspective with a professional population, whereas the majority of healthcare DCEs target a patient or general public population,[1,4] and it is possible to elicit preferences from many different perspectives.[48,49] While in principle incorporating BWS-1 into DCE attribute development should be possible in other contexts, a single study can only demonstrate feasibility in one. We discuss first methodological lessons from the current context, then potential issues which might arise in other areas.

In the current study, BWS-1 and DCE decisions were made in different contexts. The BWS-1 measured the relative importance of child and AAC device attributes in general, i.e. averaged over the case mix AAC professionals encounter in practice. In contrast, the DCE examined their decision-making for a specific child vignette.[*] Having different perspectives had benefits, in that each addressed the research question of how AAC professionals make decisions from different angles. The DCE examined trade-offs made between AAC system attributes for a child with specific characteristics. However, the set of child vignettes shown to DCE participants was not representative of the population of children seen by AAC professionals. The BWS-1 measured how important attributes were in typical practice, which the DCE could not. The disadvantage of different perspectives in the BWS-1 and DCE was a greater risk that BWS-1 RIS would be misleading as to attributes' importance in a DCE.

A different perspective in the DCE was essential in this area due to the large amount of heterogeneity in the population of children who would benefit from AAC. In areas outside AAC this may not be necessary. For example, DCEs may be used to investigate clinician preferences for screening programmes.[50,51] In that case, it would be straightforward to have the same perspective by asking in a BWS-1 what factors are considered most and least important for screening programmes, then in a DCE asking to choose which of two programmes to implement. (This perspective is classified as proxy ex-ante by Tsuchiya and Watson.[49]) Patient preferences and/or willingness-to-pay for treatments it would also be natural to have the same perspective for a BWS-1 (what is the most/least important aspect of treatment?) and a DCE (what is your preferred treatment?). (This perspective is classified as personal ex-post by Tsuchiya and Watson.[49])

As perspective may influence how participants respond to surveys[48,52-54] a possible advantage of using the same perspective for BWS-1 and DCE is greater correspondence between the participants' responses in the two formats. This may be especially important if the research is eliciting WTP. In this case, external validity could be compromised if a highly valued attribute is omitted from the DCE. Having the same perspective would also enable

---

[*] These correspond to the social ex ante case 1 and proxy ex post perspectives respectively in Tsuchiya and Watson's[49] classification system.

the estimation of joint models with BWS-1 and DCE data[33], although care should be taken in interpreting the results of such models given the methods' different properties.[34]

There are also drawbacks to using the same perspective in BWS-1 and DCE. Using different perspectives gave complementary findings, giving insight both into AAC professionals' priorities over the population of children they see in practice, as well as the detailed trade-offs made between AAC attributes when deciding for a specific child.

Depending on a study's context and target population, there may be a choice between using the same or differing perspectives in BWS-1 and DCE. In such cases, researchers should carefully consider the trade-offs between potentially greater correspondence between BWS-1 RIS and DCE attribute importance, and being able to address research questions from different perspectives. Even where there is little freedom to choose different perspectives, it is still recommended to consider the impact of participants' perspectives on the interpretation of results.

As well as theoretical considerations, this study highlighted practical issues in incorporating a BWS-1 into DCE attribute selection. First, some necessary and desirable changes to terminology were made in translating BWS-1 attributes to DCE attributes. The shift from AAC device to AAC system was discussed above, and changes in attributes' wording are shown in Table 1. Changing terminology will to some extent have altered the constructs that attributes were capturing. However, the different perspectives discussed above meant BWS-1 and DCE attributes would inevitably be conceptually different.

The experience of this study was that language and terminology was important to consider and refine throughout the project. A recommendation for future studies is to carefully define key terminology at the project's inception, and to continuously reassess and refine it as the project progresses.

Many DCEs gather information on attributes' relative importance from existing DCEs in the area. However, in AAC, no stated preference studies existed prior to this project, making BWS-1 results especially valuable. Our experience was that having BWS-1 results available in DCE attribute selection was extremely helpful. On this basis, it is recommended that incorporating BWS-1 results in DCE attribute selection will be most useful in similar areas with many potential attributes and little or no previous stated preference work.

Even in fields with several prior DCEs, BWS-1 results could still be useful, as there may be a risk of "path dependency" in the literature. If the first stated preference study in a given field omits attributes, subsequent studies are more likely to omit them as well, leading to only a narrow range of attributes being examined. For example, Janssen et al.[13] found only 10 distinct attribute categories in the area of type 2 diabetes, and Webb et al.[55] found 13 in the field of multiple sclerosis. In addition, new clinical and social developments may decrease the current relevance of past studies.

Adding a BWS-1 to DCE instrument development means a greater workload and longer project timelines than a DCE alone. In the current project, it was an advantage that it examined professionals' views, so several members of the research team were part of the target population. If patient preferences were being elicited, then stakeholder involvement would have been more resource intensive, both in developing BWS-1 attributes and translating them to DCE attributes. For example, when creating levels for DCE attributes, the research team was able to verify they

were meaningful and in an appropriate range. With a patient population this may require greater resources, for example focus groups to obtain patients' input.

It is a disadvantage of the current study that it presents only one case study in one clinical area. It also targets a professional population, whereas DCEs with patient respondents are more common.[4] Future research could usefully investigate the feasibility of the method with patients.

The population of AAC professionals who work with children in the UK is small, and many will have responded to both surveys (though data is not available on how many). It is not desirable to have DCE respondents who previously may have completed the BWS-1, as it is not known whether and in what way the earlier survey may have influenced their DCE choices. Thus a disadvantage of this method is that with a small potential respondent pool, there will inevitably be an overlap in participation between the two surveys.

5. Conclusion

We have demonstrated that a BWS-1 can be useful in informing attribute selection for a DCE. This study has also shown that using BWS-1 and DCE as part of a pre-defined research project can produce synergies, giving greater insight than carrying out each study singly, especially in a complex decision-making environment where no prior stated preference work has been conducted.

We propose that incorporating a BWS-1 into the process of DCE instrument development is a valid and robust method, and we finish with a set of recommendations for future studies:-

1) Have a clear plan of how DCE attribute selection will proceed, including what selection criteria will be used in addition to BWS-1 results. For example, researchers may establish a priori that to address the research question some attributes will be included in the DCE regardless of BWS-1 results.

2) Clearly define the decision-making context and participants' perspectives in both the BWS-1 and DCE, and consider the impact perspective will have on findings.

3) Clearly define key terminology at the beginning of the project and refine it as necessary as it progresses in the light of interim findings.

4) Consider using a BWS-1 to inform DCE attribute development and selection, especially in areas with little prior stated preference work and/or extensive qualitative work is difficult. However, there may still be many advantages to the approach where this is not the case.

References

1. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. PharmacoEconomics 2014;32:883-902.

2. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Economics 2012;21:145-72.

3. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied health economics and health policy 2003;2:55-64.

4. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019;37:201-26.

5. Bridges JF, Hauber AB, Marshall D, 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;14:403-13.

6. dosReis S, Castillo WC, Ross M, Fitz-Randolph M, Vaughn-Lee A, Butler B. Attribute development using continuous stakeholder engagement to prioritize treatment decisions: a framework for patient-centered research. Value in Health 2016;19:758-66.

7. Hollin IL, Young C, Hanson C, Bridges JF, Peay H. Developing a patient-centered benefit-risk survey: a community-engaged process. Value in Health 2016;19:751-7.

8. Michaels-Igbokwe C, Lagarde M, Cairns J, Terris-Prestholt F. Using decision mapping to inform the development of a stated choice survey to elicit youth preferences for sexual and reproductive health and HIV services in rural Malawi. Social Science & Medicine 2014;105:93-102.

9. Kløjgaard ME, Bech M, Søgaard R. Designing a stated choice experiment: the value of a qualitative process. Journal of Choice Modelling 2012;5:1-18.

10. Coast J, Al‐Janabi H, Sutton EJ, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health economics 2012;21:730-41.

11. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of health services research & policy 2007;12:25-30.

12. Wittenberg E. Instrument Development in Choice Experiments. Commentary on:"Applying a Framework for Instrument Development of a Choice Experiment to Measure Treatment Preferences in Type 2 Diabetes". The Patient-Patient-Centered Outcomes Research 2016;9:379-81.

13. Janssen EM, Segal JB, Bridges JFP. A Framework for Instrument Development of a Choice Experiment: An Application to Type 2 Diabetes. The Patient - Patient-Centered Outcomes Research 2016;9:465-79.

14.   Seo J, Smith BD, Estey E, Voyard E, O' Donoghue B, Bridges JFP. Developing an instrument to assess patient preferences for benefits and risks of treating acute myeloid leukemia to promote patient-focused drug development. Current Medical Research and Opinion 2018;34:2031-9.

15.   Bridges JFP, Oakes AH, Reinhart CA, Voyard E, O'Donoghue B. Developing and piloting an instrument to prioritize the worries of patients with acute myeloid leukemia. Patient preference and adherence 2018;12:647-55.

16.   Beusterien K, Chan E, Such P, et al. Development of a stated-preference instrument to prioritize treatment goals in recent onset schizophrenia. Current Medical Research and Opinion 2017;33:2129-36.

17.   Katz DA, Stewart KR, Paez M, et al. Development of a Discrete Choice Experiment (DCE) Questionnaire to Understand Veterans' Preferences for Tobacco Treatment in Primary Care. The Patient - Patient-Centered Outcomes Research 2018;11:649-63.

18.   Barber S, Bekker H, Marti J, Pavitt S, Khambay B, Meads D. Development of a Discrete-Choice Experiment (DCE) to Elicit Adolescent and Parent Preferences for Hypodontia Treatment. The Patient - Patient-Centered Outcomes Research 2019;12:137-48.

19.   Webb EJ, Meads D, Lynch Y, et al. What's important in AAC decision making for children? Evidence from a best–worst scaling survey. Augmentative and Alternative Communication 2019:1-15.

20.   Webb EJ, Lynch Y, Meads D, et al. Finding the best fit: examining the decision-making of augmentative and alternative communication professionals in the UK using a discrete choice experiment. BMJ open 2019;9.

21.   Cheung KL, Wijnen BF, Hollin IL, et al. Using Best-Worst Scaling to Investigate Preferences in Health Care. Pharmacoeconomics 2016;34:1195-209.

22.   Hofstede SN, van Bodegom-Vos L, Wentink MM, et al. Most important factors for the implementation of shared decision making in sciatica care: ranking among professionals and patients. PLoS One 2014;9:e94176.

23.   Whitty JA, Gonçalves ASO. A systematic review comparing the acceptability, validity and concordance of discrete choice experiments and best–worst scaling for eliciting preferences in healthcare. The Patient-Patient-Centered Outcomes Research 2018;11:301-17.

24.   Krucien N, Watson V, Ryan M. Is best–worst scaling suitable for health state valuation? A comparison with discrete choice experiments. Health economics 2017;26:e1-e16.

25.   Flynn TN, Peters TJ, Coast J. Quantifying response shift or adaptation effects in quality of life by synthesising best-worst scaling and discrete choice data. Journal of choice modelling 2013;6:34-43.

26.   Whitty JA, Ratcliffe J, Chen G, Scuffham PA. Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best-worst scaling methods. Medical decision making : an international journal of the Society for Medical Decision Making 2014;34:638-54.

27.   Whitty JA, Walker R, Golenko X, Ratcliffe J. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. PloS one 2014;9:e90635.

28.   Potoglou D, Burge P, Flynn T, et al. Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc Sci Med 2011;72:1717-27.

29.   Severin F, Schmidtke J, Mühlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. European Journal of Human Genetics 2013;21:1202-8.

30.   Mansfield C, Ndife B, Chen J, Gallaher K, Ghate S. Patient preferences for treatment of metastatic melanoma. Future Oncol 2019;15:1255-68.

31.   Marshall T, Pugh A, Fairchild A, Hass S. Patient Preferences for Device-Aided Treatments Indicated for Advanced Parkinson Disease. Value Health 2017;20:1383-93.

32.   Turk D, Boeri M, Abraham L, et al. Patient Preferences for Osteoarthritis Pain and Chronic Low Back Pain Treatments in the United States: A Discrete-Choice Experiment. Osteoarthritis and cartilage 2020;08.

33.   Zhang J, Johnson FR, Mohamed AF, Hauber AB. Too many attributes: a test of the validity of combining discrete-choice and best–worst scaling data. Journal of choice modelling 2015;15:1-13.

34.   Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP capability index for older people. Social science & Medicine 2008;67:874-82.

35.   Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. Journal of health economics 2007;26:171-89.

36.   Murray J, Lynch Y, Goldbart J, et al. Identifying appropriate communication aids for children and public involvement in research: a mixed methods studyIn press. Report No.: 14/70/153.

37.   Beukelman DR, Mirenda P. Augmentative and alternative communication: Supporting children and adults with complex communication needs: Paul H. Brookes Pub.; 2013.

38.   Murray J, Lynch Y, Meredith S, et al. Professionals' decision-making in recommending communication aids in the UK: competing considerations. 2019:1-13.

39.   Lynch Y, Murray J, Moulam L, et al. Decision-making in communication aid recommendations in the UK: cultural and contextual influencers. 2019:1-13.

40. Judge S, Randall N, Goldbart J, et al. The language and communication attributes of graphic symbol communication aids–a systematic review and narrative synthesis. 2019:1-11.

41. Thistle JJ, Wilkinson KM. Building evidence-based practice in AAC display design for young children: Current practices and future directions. Augmentative and Alternative Communication 2015;31:124-36.

42. Dada S, Alant E. The effect of aided language stimulation on vocabulary acquisition in children with little or no functional speech. American Journal of Speech-Language Pathology 2009;18:50-64.

43. Janssen EM, Bridges JF. Art and science of instrument development for stated-preference methods. Springer; 2017.

44. Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. Pharmacoeconomics 2017;35:697-716.

45. Armatas CA, Venn TJ, Watson AE. Applying Q-methodology to select and define attributes for non-market valuation: A case study from Northwest Wyoming, United States. Ecological Economics 2014;107:447-56.

46. Kerr GN, Swaffield SR. Identifying cultural service values of a small river in the agricultural landscape of Canterbury, New Zealand, using combined methods. Society & Natural Resources 2012;25:1330-9.

47. Louviere JJ, Street D, Burgess L, Wasi N, Islam T, Marley AA. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. Journal of choice modelling 2008;1:128-64.

48. Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. Health economics 2003;12:545-51.

49. Tsuchiya A, Watson V. Re‐Thinking 'the different perspectives that can be used when eliciting preferences in health'. Health economics 2017;26:e103-e7.

50. Sicsic J, Krucien N, Franc C. What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. Social Science & Medicine 2016;167:116-27.

51. Boone D, Mallett S, Zhu S, et al. Patients'& healthcare professionals' values regarding true-& false-positive diagnosis when colorectal cancer screening by CT colonography: discrete choice experiment. PloS one 2013;8:e80767.

52. Damschroder LJ, Roberts TR, Zikmund-Fisher BJ, Ubel PA. Why people refuse to make tradeoffs in person tradeoff elicitations: A matter of perspective? Medical Decision Making 2007;27:266-80.

53. Bobinac A, Van Exel NJA, Rutten FF, Brouwer WB. Valuing QALY gains by applying a societal perspective. Health economics 2013;22:1272-81.

54. Gyrd-Hansen D. Willingness to pay for a QALY. Pharmacoeconomics 2005;23:423-32.

55. Webb EJD, Meads D, Eskyte I, et al. A Systematic Review of Discrete-Choice Experiments and Conjoint Analysis Studies in People with Multiple Sclerosis. The Patient - Patient-Centered Outcomes Research 2018;11:391-402.

Table 1: Best-worst scaling case 1 attributes

Child attributes	Included in DCE?
Access to professional AAC support	
Attention level	
Child's determination and persistence	Yes
Child's receptive and expressive language abilities	Yes
Communication ability with aided AAC	Yes
Educational stage	
Functional visual skills	
History of aided AAC use	
Insight into own communicative skills	
Level of fatigue	
Level of learning ability	
Literacy ability	
Mobility	
Physical abilities for access	
Predicted future needs and abilities	Yes
Presence of additional diagnoses	
Primary diagnosis	
Speech skills and intelligibility	
Support for AAC from communication partners	
AAC device attributes	
Additional assistive technology functions	
Appearance	
Battery life	
Consistency of layout and navigation	Yes
Cost	
Durability and reliability	
Ease of customization	
Ease of mounting on a range of equipment	
Graphic representation	Yes
Number of cells per page	
Number of key presses required to generate symbol or text output	
Portability	
Range of access methods	
Size of output vocabulary	Yes
Supplier support	
Type of vocabulary organization	Yes
Vocabulary or language package(s)	Yes
Voice	

Note. DCE = discrete choice experiment

Table 2: Discrete choice experiment attributes and levels and corresponding best-worst scaling case 1 attributes

BWS attribute	BWS rank	DCE attribute	DCE levels
Child attributes			
Child's receptive and expressive language abilities	1	Receptive and expressive language	Delayed Receptive language exceeding expressive language
Communication ability with aided AAC	3	Communication ability with AAC	No previous AAC experience Abe to use AAC for a few communicative functions Able to use AAC for a range of communicative functions
Child's determination and persistence	4	Child's determination and persistence	Does not appear motivated to communicate through any methods and means Motivated to communicate through symbol communication systems Only motivated to communicate through methods other than symbol communication
Predicted future needs and abilities	6	Predicted future skills and abilities	Regression Plateau Progression
AAC device/system attributes			
Vocabulary or language package(s)	1	Vocabulary sets	No vocabulary set Fixed vocabulary set Vocabulary set with staged progression
Consistency of layout and navigation	2	Consistency of layout	Consistency of some aspects of layout Consistency of all aspects of layout Idiosyncratic layout
Type of vocabulary organization	5	Type of vocabulary organisation	Visual scene Taxonomic Semantic-syntactic Pragmatic
Size of output vocabulary	7	Size of vocabulary	Up to 50 vocabulary items 50-1000 vocabulary items More than 1000 vocabulary items
Graphic representation	12	Graphic representation	Photos Pictographic symbol set Ideographic symbol system (with rules or encoding) Text

Note. DCE = discrete choice experiment; BWS = best-worst scaling

DCE attribute selection incorporating a BWS survey

Figure 1: Timeline of project events

BWS – Best-worst scaling

DCE – Discrete choice experiment

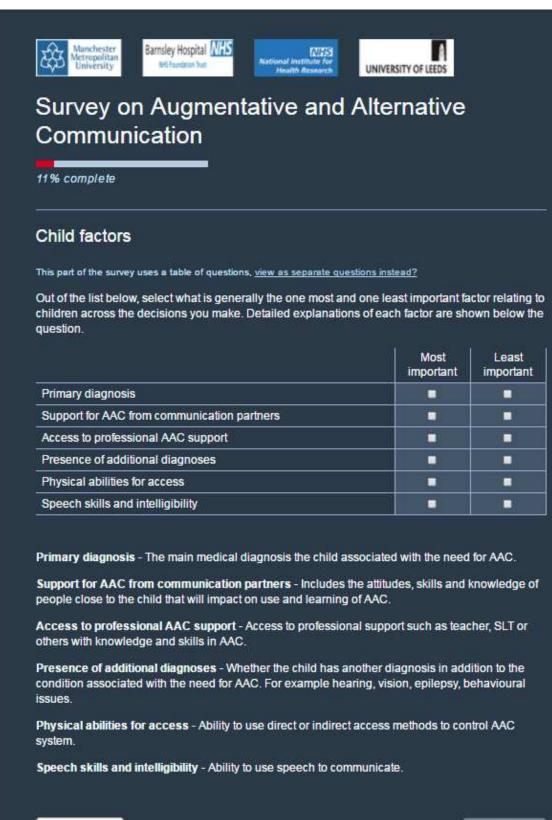**Figure 2: Example best-worst scaling task**

Figure 3: Example discrete choice experiment task

Child B has delayed expressive and receptive language and able to use aided AAC for a few communicative functions. Child B is only motivated to communicate through methods other than symbol communication systems. Child B is predicted to maintain current skills and abilities (plateau).

	System 1	System 2	System 3
Graphic Representation Primary type of graphic symbol used	Ideographic symbols	Photo symbols	Pictographic symbols
Consistency of layout Consistency of layout of symbols on pages, including when navigating through pages to select desired output.	Highly consistent layout	Somewhat consistent layout	Inconsistent layout
Vocabulary sets Pre-determined vocabulary or language package provided	Commercially provided sets without language progression	Commercially provided sets with language progression	Commercially provided sets without language progression
Size of vocabulary The size of the output vocabulary available within the aided AAC system.	More than 1000 vocabulary items	50-1000 vocabulary items	Up to 50 vocabulary items
Type of vocabulary organisation Primary format used to organise the vocabulary within the aided AAC system	Pragmatic organisation	Semantic syntactic organisation	Visual scene display

For this child I would choose: ○   ○   ○

On a scale from 1 to 7, how good a match is your chosen device for this child? (1=very unsuitable, 7=very suitable)

1 = very unsuitable	2	3	4	5	6	7 = very suitable
○	○	○	○	○	○	○

← →

35%