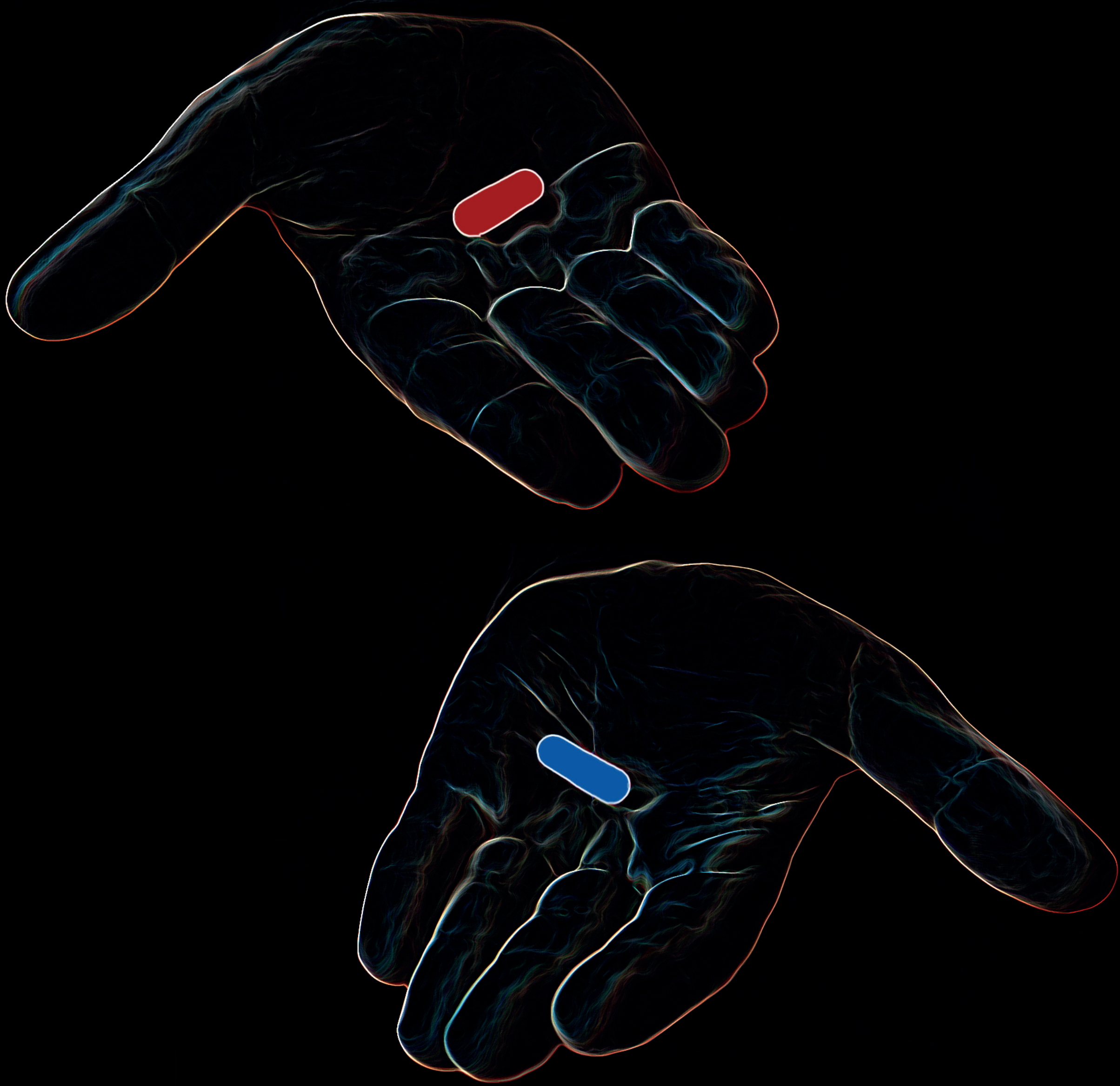
Choice Modelling in Health: Challenges and Opportunities

Vikas Rogier Soekhai



Choice Modelling in Health: Challenges and Opportunities

Vikas Rogier Soekhai

ISBN: 978-94-9042-065-9

Cover design: Vikas Rogier Soekhai and Ipskamp Printing
Lay-out: Vikas Rogier Soekhai
Printing: Ipskamp Printing

Choice Modelling in Health: Challenges and Opportunities

Keuzemodellering in de gezondheidszorg: uitdagingen en kansen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 4 november 2021 om 10.30 uur

door

Vikas Rogier Soekhai
geboren te Leiden

Erasmus University Rotterdam

Promotiecommissie

Promotoren: Prof.dr. E.W. de Bekker-Grob
Prof.dr. B. Donkers

Overige leden: Prof.dr. M.P.M.H. Rutten - van Mölken
Prof.dr.ir. C.G. Chorus
Dr. C.G.M. Oudshoorn

Contents

Chapter 1

General Introduction

1.1 Background of choice modelling in health

The field of choice modelling is concerned with understanding how individuals make choices by quantifying the underlying preferences.[1] More specifically, choice modelling aims to characterize choices individuals make (e.g., what drives an individual's choice for alternative A or B) and to predict choices among the alternatives considered. In economic and psychology literature the concepts of preferences and utilities are considered the best notion of why an individual selects alternative A over B.[1] In choice modelling it is assumed that choice behavior is based on preferences (often called taste), determining the amount of pleasure or satisfaction they derive from goods and services.[2] This satisfaction is called utility and preferences are quantified by analyzing the utility functions of choice alternatives. These alternatives can be decomposed into characteristics (attributes) describing them, which contribute to the total utility of choice alternatives. Choice modelling relates individuals' choices to their preferences by focusing on the utilities of the alternatives available.

Choice modelling (i.e., specifically discrete choice experiments, see section 1.2) in health was introduced in the early 1990s, especially enhancing health benefit assessments as a way to not only capture health outcomes but also outcomes that extend beyond health.[3] Before the introduction of choice modelling in health, methods from other fields (e.g., contingent valuation and time trade-off) were used to gain insights into health preferences.[4,5] Using choice modelling to gain insights into these preferences provided several advantages over the methods used in that time period.[3,6] First, choice modelling enabled researchers to collect high-quality preference data at a lower cost. Second, it provided information about the incremental benefit of the characteristics of the choice alternative of interest. Last, it allowed a way to overcome methodological issues in gaining insights into health preferences using for example contingent valuation. After the 1990s the interest in the application of choice modelling in health continued to grow, with not only applications on the macro - level (health benefit assessments to inform policy-making) but also on micro-level (patient preferences for shared decision-making).[7]

Health preferences can be explored or elicited through revealed – or stated-preference methods. Revealed-preference methods are based on actual real-life behavior, while stated-preference methods are considering hypothetical situations in a controlled setting.[1,8] In this dissertation we focus on stated-preferences methods since they are more widely used in health because real-life data is often not available.[9,10] Within stated-preference methods we can distinguish between methods allowing us to explore (exploration methods, often qualitative) and elicit (elicitation methods, often quantitative) preferences, since there are several ways of gaining insights into preferences. There have been several studies providing an overview of preference methods in health, but an up-to-date compendium of methods to either explore or elicit preferences can serve as an important resource to assess these methods and is important to further drive research on the incorporation of preferences in health decision-making forward.[11,12]

1.2 Discrete choice experiments

Among the preference elicitation methods, discrete choice experiments (DCEs) are increasingly advocated.[13–15] A DCE is a survey-based preference elicitation method in which individuals are asked to select their preferred alternative from a set of alternatives. Individuals select only one alternative from the set of alternatives. The DCE method is often used to study preferences for interventions, products or services for which no market yet exists, for example to gain insights into patients' acceptance of side effects for a new drug treatment. A DCE survey consists of a number of choice situations (choice tasks) in which individuals are asked to select their preferred choice from the presented alternatives. These alternatives can be decomposed into characteristics (attributes) describing them. In each choice task, alternatives differ from each other by systematic variation in attribute values (attribute levels). Figure 1 presents an example DCE choice task.

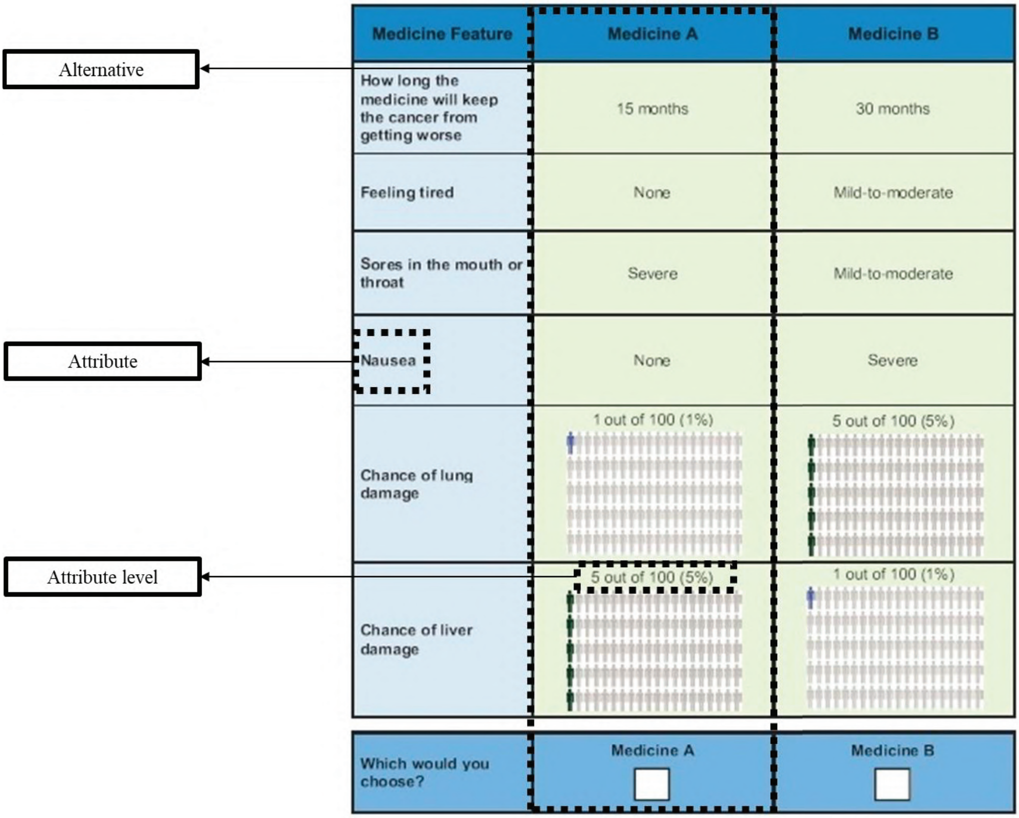


Figure 1 Example DCE choice task[16]

DCE data analysis has its origin in mathematical psychology, with wide applications in marketing, transport and environmental economics.[1,17–19] The body of academic DCE literature increased rapidly after its introduction in health in the 1990s.[15,20,21] Although there have been several reviews investigating the way DCE studies build up in the literature, it is important to keep track of developments of DCEs in health to examine whether challenges identified in prior reviews are still relevant or if there has been a response to published suggestions and guidelines.[14,15,22]

The theoretical foundation of DCEs is in random utility theory (RUT), described by Thurstone and eventually extended by Nobel Prizewinner McFadden.[23,24] RUT is also consistent with Lancaster's theory of value, in which it is assumed that alternatives can be valued in terms of their characteristics.[25] Based on these theories, in DCEs it is assumed that (1) alternatives can be described by their attributes, (2) an individual's valuation depends upon the levels of these attributes and (3) choices are based on a specific utility function. Within the RUT framework, utility maximization is assumed when modelling DCE data.[1,26] Furthermore, the RUT framework states that utility (U) can be partitioned into a systematic part (V) that is driven by individuals' stable preferences, which the analyst can capture. Second, an unobserved residual component (ε) representing the part of utility that cannot be captured by the analyst (unobserved utility component). Considering the analyst only observes choices and not the underlying true utility levels, probabilistic models are used to account for the unobserved utility component when analyzing DCE data.[26] This leads to a probability, in the situation of alternatives a and b for example, of selecting alternative a over alternative b when:

$$P(Y = a) = P(V_a + \varepsilon_a > V_b + \varepsilon_b) \qquad \text{eq. 1}$$

Meaning that an individual selects the alternative with the highest utility. The systematic utility part can be specified for alternative j as:

$$V_a = \beta' X_a \qquad \text{eq.2}$$

where X_a represents a vector of observed attributes relating to alternative a and β the coefficients for these attributes.[26] This means that the impact of attributes on V is estimated based on choices of the individual in de DCE survey. The β coefficients for an attribute represent the utility weight of that attribute on the total utility of the alternative. Larger coefficients[i] represent more importance regarding the individual's choice for that attribute compared to other attributes. These coefficients can be used to calculate the relative importance of attributes, trade-offs between attributes and to predict shares of choices (choice shares).

DCEs are conducted following specific academic guidelines and consist of following different steps.[27,28] First, literature study and qualitative research is used to select attributes and levels, after which decisions on the choice task are made. Based on the utility function the analyst wants to estimate, the experimental design is constructed. This design includes specific combinations of attributes and levels to be used in the DCE, leading to the estimation of unbiased coefficients. Finally, the DCE survey is developed and data-collection is started using a pilot study to detect errors and update the experimental design. Once final data is collected, econometric modelling is used to estimate coefficients for the attributes.

1.3 Case 2 Best-Worst Scaling

Next to DCEs, best-worst scaling (BWS) has become an increasingly popular method to elicit preferences in health and healthcare, especially for health state and medical treatment valuation.[29,30] The introduction of BWS came from the intent to obtain more preference information than from a discrete choice experiment (DCE) by asking individuals to select their "best" and "worst" option, without increasing the cognitive burden.[31,32] There are several types of BWS, but case 2 BWS (BWS-2), or often referred to as profile case BWS, received

[i] Because of differences in utility scales, coefficients are only comparable in absolute terms within a model, not between models

much attention in health economics as this method can uncover attribute level importance, reduce cognitive burden of the choice task by focusing on one profile at a time and are relatively easy to design.[33,34] See Figure 2 for an example BWS-2 choice task.

The theoretical foundation and model-based estimation in BWS-2 is also based on the RUT framework, similar to DCEs. However, we do need to highlight two aspects here. Firstly, in BWS-2 individuals make choices for "best" and "worst" on the attribute level instead of the level of an alternative in DCEs. This means that in the situation of two attributes for example, selecting attribute a with level y over attribute b with level z given by:

$$P(Y_{best} = Att_{a,y}) = P\left(V_{a,y} + \varepsilon_{a,y} > V_{b,z} + \varepsilon_{b,z}\right) \qquad \text{eq. 3}$$

Secondly, the analyst can assume two psychological processes of decision-making when analyzing BWS-2 choice data: (1) the maximum difference model (maxdiff), in which

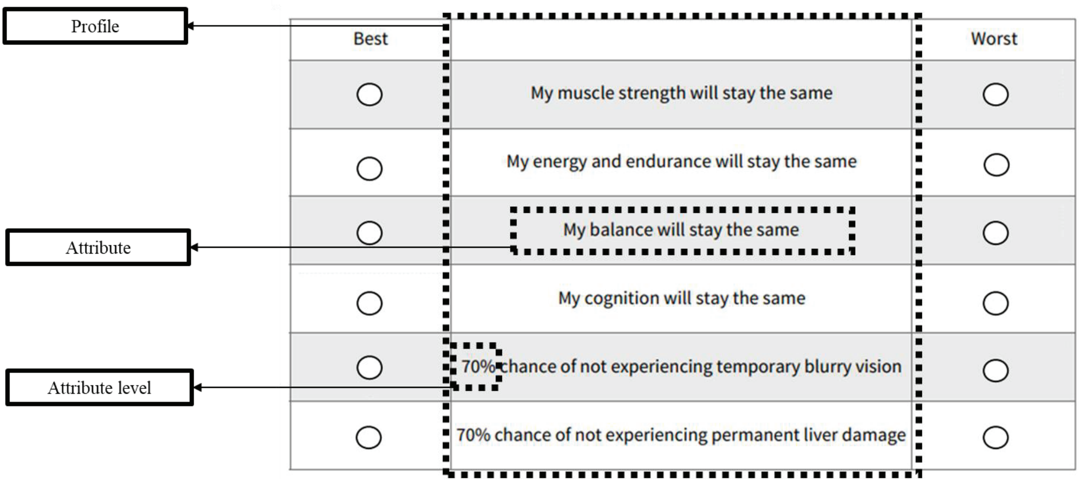


Figure 2 Example BWS-2 choice task

individuals select that best-worst pair that maximizes the utility difference between "best" and "worst", and (2) the sequential model in which individuals make their "best" and "worst" choices in two stages: first choosing the “best” (“worst”) from all options and then selecting the “worst” (“best”) from all remaining options.[31]

Although there are no specific academic guidelines on how to conduct BWS-2, the process is similar to DCEs: from identification of attributes and levels to generating the experimental design and econometric modelling of the BWS-2 choice data. A resulting benefit of BWS-2 in contrast to DCEs, is that it is able to provide an overall ranking of attribute levels since every level is measured on the same utility scale with the same reference level.[29]

1.4 Methodological challenges and opportunities

Both DCE and BWS-2 have become more popular preference methods in health, although there are also disadvantages associated with both methods. DCEs can for example be complex and therefore cognitively demanding, while in BWS-2 individuals are always forced to make a choice.[1,31] To ensure that these methods in general become more valuable for actual decision-making (i.e., at the policy or clinical level), several methodological challenges need to be overcome. In DCEs it is for example interesting to study which econometric modelling techniques and analyses provide the most accurate policy-relevant outcomes, like for example choice share predictions. For BWS-2, several issues relating to its design and analysis require further exposition. One of these issues is the inclusion of a mixture of positive (e.g., treatment effectiveness) and negative (e.g., treatment side effects) attributes in BWS-2 tasks and its effect on the estimation of coefficients, how attribute framing[ii] impacts estimates and the role of reference points in BWS-2. Furthermore, it is also important, given the rise of DCE and BWS-2 in healthcare applications, to compare outcomes between these two methods. Not only in terms of estimated coefficients and related outcomes, but also in terms of cognitive

[ii] Turning a positive (chance of being cured) attribute into a negative (chance of not being cured).

burden for example. Gaining insights into methodological challenges and providing opportunities to overcome them, contributes to academic literature as well as practice to inform decision-making.

1.5 Dissertation objectives

The application of both DCE and BWS-2 in health to inform decision-making seems promising. This dissertation addresses methodological challenges and opportunities as well as practical applications of both methods. Main objectives are:

1. **Providing insights into preference methods used in health**
    a. An overview of current preference methods used in health
    b. An overview of applications and methods by DCEs in health

2. **Providing insights into DCE and BWS-2 challenges and opportunities regarding design and analysis**
    a. Study whether choice share predictions in DCE depend on modelling and analysis approach
    b. Study the impact of mixing positive and negative attributes in BWS-2
    c. Study the impact of attribute framing in BWS-2
    d. Study the impact of including explicit reference points in BWS-2

3. **Empirically comparing outcomes between DCE and BWS-2**
    a. Study differences in perceived cognitive burden between DCE and BWS-2
    b. Study differences in statistical outcomes and policy relevant measures between DCE and BWS-2

1.6 Dissertation structure

This thesis consists of three parts. The first part (**chapters 2-3)** provides an overview of health preference methods used in health economics and more specifically focuses on an overview of current DCE practice. Both studies consisted of a systematic literature review. These chapters provide insights in current trends in health preference research in general.

Part two consists of methodological challenges and opportunities for both DCE and BWS-2. To study whether the accuracy of choice share predictions in DCEs depends on the modelling approach[iii] being used as well as the type of analysis being conducted to calculate these predictions, a simulation study was conducted (**chapter 4)**. **Chapter 5** focuses on the effect of mixing positive and negative attributes in BWS-2 tasks on estimated coefficients, which is illustrated both analytically and with simulation examples. **Chapter 6** describes an empirical study investigating the impact of framing attributes in BWS-2 on estimation outcomes, while **chapter 7** outlines both analytically as well as empirically the role of reference points in BWS-2 tasks.

Part three reports the outcomes of empirical studies focused on the differences in perceived cognitive burden of DCE and BWS-2 (**chapter 8**), as well as differences in statistical outcomes between these two methods (**chapter 9**). This dissertation ends with a general discussion in which the outcomes of the previous chapters are integrated and further discussed. This section also includes a number of conclusions and recommendations for future research and policy-making.

[iii] Econometric models that do and do not account for variation in preferences across individuals.

References

1. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*. Cambridge University Press; 2015. doi:10.1007/9781316136232

2. Reisch LA, Zhao M. Behavioural Economics, Consumer Behaviour, and Consumer Policy: State of the Art. *Behav Public Policy*. 2017;1(2):190–206. doi:10.1017/BPP.2017.1

3. Ryan M, Gerard K, Amaya-Amaya M. *Using Discrete Choice Experiments to Value Health and Health Care*. Springer; 2008.

4. Pullman ME, Dodson KJ, Moore WL. A Comparison of Conjoint Methods When There Are Many Attributes. *Mark Lett 999 102*. 1999;10(2):125-138. doi:10.1023/A:1008036829555

5. Louviere J. Analyzing Decision Making: Metric Conjoint Analysis. In: ; 1988.

6. Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. *Health Econ*. 2003;12(1):3-16. doi:10.1002/HEC.763

7. Single ANV, Facey KM, Livingstone H, Silva AS. Stories of Patient Involvement Impact in Health Technology Assessments: A Discussion Paper. *Int J Technol Assess Health Care*. 2019;35(4):266-272. doi:10.1017/S0266462319000552

8. Mark T, Swait J. Using stated preference and revealed preference modeling to evaluate prescribing decisions. *Health Econ*. 2004;13(6):563-573. doi:10.1002/HEC.845

9. Lambooij MS, Harmsen IA, Veldwijk J, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC Med Res Methodol 2015 151*. 2015;15(1):1-8. doi:10.1186/S12874-015-0010-5

10. Smith H, Currie C, Chaiwuttisak P, Kyprianou, A. Patient choice modelling: how do patients choose their hospitals? *Health Care Manag Sci*. 2018;21(2):259-268. doi:10.1007/S10729-017-9399-1

11. Ryan, M R, Scott, DA S, Reeves, C R, et al. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*. 2001;5(5). doi:10.3310/HTA5050

12. (MDIC) MDIC. MEDICAL DEVICE INNOVATION CONSORTIUM (MDIC) PATIENT CENTERED BENEFIT-RISK PROJECT REPORT: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology By Medical Device I. Accessed July 8, 2018. http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf

13. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64. Accessed June 18, 2018. http://www.ncbi.nlm.nih.gov/pubmed/14619274

14. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145-172. doi:10.1002/hec.1697

15. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics*. 2014;32(9):883-902. doi:10.1007/s40273-014-0170-x

16. Adapted from: Discrete Choice Experiments and Conjoint Analysis. Accessed July 13, 2021. https://www.rtihs.org/services/health-preference-assessment/discrete-choice-experiments-and-conjoint-analysis

17. Hensher DA. Stated preference analysis of travel choices: the state of practice. *Transp 1994 212*. 1994;21(2):107-133. doi:10.1007/BF01098788

18. Adamowicz V, Boxall P. Future Directions of Stated Choice Methods for Environment Valuation. *undefined*. Published online 2001.

19. Mahieu P-A, Andersson H, Beaumais O, Crastes dit Sourd R, Hess S, Wolff F-C. Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. *Rev Agric Food Environ Stud*. 2017;98(3):201-220. doi:10.1007/S41130-017-

0053-6

20. Propper C. Contingent Valuation of Time Spent on NHS Waiting Lists. *Econ J*. 1990;100(400):193-199. doi:10.2307/2234196

21. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. *Health Econ*. 1997;6(3):261-273. doi:10.1002/(sici)1099-1050(199705)6:3<261::aid-hec262>3.0.co;2-n

22. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64. Accessed July 13, 2021. https://europepmc.org/article/med/14619274

23. Thurstone LL. A law of comparative judgment. *Psychol Rev*. 1927;34(4):273-286. doi:10.1037/h0070288

24. McFadden D. Conditional logit analysis of qualitative choice behavior. *Front Econom*. Published online 1974.

25. Lancaster KJ. A New Approach to Consumer Theory. *J Polit Econ*. 1966;74(2):132-157. doi:10.1086/259131

26. Train K. *Discrete Choice Methods with Simulation*. Cambridge University Press; 2002.

27. Bridges JFP, 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 Heal*. 2011;14(4):403-413. doi:10.1016/j.jval.2010.11.013

28. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Health Care Decision Making : A User ' s Guide , Experiments to Inform Healthcare A User ' s Guide. *Pharmacoeconomics*. 2008;26(December 2015):661-677. doi:10.2165/00019053-200826080-00004

29. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-189. doi:10.1016/j.jhealeco.2006.04.002

30. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):1-14. doi:10.1186/s13561-015-0079-x

31. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*.; 2015. doi:10.1017/CBO9781107337855

32. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):259-267. doi:10.1586/erp.10.29

33. van Dijk JD, Groothuis-Oudshoorn CGM, Marshall DA, IJzerman MJ. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. *Value Heal*. 2016;19(4):316-322. doi:10.1016/j.jval.2015.12.020

34. 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. *Med Decis Mak*. 2014;34(5):638-654. doi:10.1177/0272989X14526640

Part I

Insights into preference methods in health

Chapter 2

Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review

Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto C, Donkers B, Huys I, van Overbeeke E, Juhaeri J and de Bekker-Grob EW

Published Drug Discovery Today 2019

Abstract

Preference studies are becoming increasingly important within the medical product decision-making context. Currently, there is limited understanding of the range of methods to gain insights into patient preferences. We developed a compendium and taxonomy of preference exploration (qualitative) and elicitation (quantitative) methods by conducting a systematic literature review to identify these methods. This review was followed by analyzing prior preference method reviews to cross-validate our results and consulting intercontinental experts to confirm our outcomes. This resulted in the identification of 32 unique preference methods. The developed compendium and taxonomy can serve as an important resource for assessing these methods and helping to determine which are most appropriate for different research questions at varying points in the medical product lifecycle.

2.1 Introduction

There is an emerging consensus that the patient perspective should be incorporated within decisions in the medical product lifecycle (MPLC) [1-4], where the medical product lifecycle in this study is defined as the lifecycles of drugs, biologics and medical devices. Broadly encouraging the involvement of patients has, therefore, become increasingly important [5,6]. Taking the patient voice into consideration has not only become increasingly important for companies that develop new medical products, but also for the authorities that assess, regulate, and decide which products are effective, safe, well-tolerated, and cost-effective [7-16].

To incorporate the patient voice, patient preferences need to be explicitly explored or elicited through revealed- or stated-preference methods. In this paper preference exploration methods are defined as qualitative methods that collect descriptive data through participant or phenomenon observation, and examining the subjective experiences and decisions made by participants. Elicitation methods are defined as quantitative methods collecting quantifiable data for hypothesis testing and other statistical analyses. While the use of revealed-preference methods still represents a methodological challenge in health, many different methods exist to assess patients' stated-preferences [17,18]. An up-to-date compendium of different stated-preference methods to explore or elicit patient preferences within the MPLC is missing.

There have been few publications on what methods can be used to assess patient preferences in a scientific, sound way in the context of the MPLC specifically. In 2001, Ryan *et al.* [19] provided an overview of methods known at the time for eliciting public preferences for healthcare. In 2015, the Medical Device Innovation Consortium (MDIC) developed an overview of different preference elicitation methods as part of their framework on incorporation of patient preferences into regulatory assessments of medical devices [20]. Although both publications made useful contributions, the study from Ryan *et al.* [19] does not reflect methods developed since 2001, and the study from the MDIC [20] did not include preference exploration methods or use a systematic approach for identifying preference elicitation methods.

Therefore, the aim of our study was to develop an up-to-date compendium and taxonomy of both exploration and elicitation preference methods within the MPLC context. This will be an important step to further drive the incorporation of patient preferences forward, in addition to the study of van Overbeeke *et al.* [6], and in developing guidance on when and how to assess patient preferences scientifically in the context of decision-making in the MPLC.

2.2 Compendium of preference methods

A systematic literature review was conducted, followed by an analysis of prior reviews by Ryan *et al.* [19] and from the MDIC [20] and expert consultations with international preference experts, to identify all potential preference exploration and elicitation methods within the context of the MPLC. In this paper, a broad definition of a preference method was used: any method that enabled us to gain insight into patients' relative desirability or acceptability of specified alternatives; or choices among treatment alternatives or outcomes; or other attributes that differ among alternative health interventions [7]. Ultimately, 208 papers were analyzed during the systematic literature review to identify preference exploration and elicitation methods within the context of the MPLC. For more information about the approach used in the systematic literature review, see Appendix A.1 of the supplementary material. An alphabetical overview of all reviewed full-text papers is listed in Appendix B of the supplementary material.

We identified 19 different methods: 5 exploration methods and 14 elicitation methods, in the systematic literature review. Most frequently cited exploration methods included focus groups (n=29, 13.9%) and (semi-) structured individual interviews (n=47, 22.6%), whereas most cited elicitation methods papers included discrete choice experiments (n=57, 27.4%) and the visual analogue scale (n=12, 5.8%). Contingent valuation (n=11, 5.3%), standard gamble (n=11, 5.3%) and time trade-off (n=11, 5.3%) were also frequently included in the analyzed papers. Four studies included best-worst scaling Type 1,2 (n=4, 2.0%).

Through the analysis of the preference methods reviews of Ryan *et al.* [19] and the MDIC [20], and after condensing several of these methods, we identified 23 preference exploration and elicitation methods. This selection included 9 preference exploration and 14 elicitation methods. From these 23 preference methods, 13 methods were also identified in our systematic literature review (56.0%). The expert consultations confirmed the methods identified in the systematic literature review and in the analysis of prior preference method reviews. Also, consensus was reached on including four additional elicitation methods. The expert consultations also resulted in the exclusion of methods focusing on scale-related (e.g. likert scales) or decision-making framework-related (e.g. multi-criteria decision analysis) techniques, since these techniques were regarded as inconsistent with our definition of a preference method.

As described above, we identified 19 methods through the systematic literature review, the 23 methods through the analysis of previously-conducted reviews, and the four additional methods via expert consultations. In total 32 unique preference methods were identified: 10 exploration and 22 elicitation methods. Table 1 summarizes and briefly describes these methods.

Table 1 Overview of identified methods

Method	Description	Reference
Exploration methods		
Citizens' juries[β]	Group of individuals discussing issues on the basis of evidence provided by two trained moderators	[24,25]
Complaints procedures[β]	Method in which stakeholders can register complaints in order to be investigated by experts	[26,27]
Concept mapping[β]	Method that utilizes small groups of participants responding to various topics or issues, while ensuring each respondent is given equal opportunity to express their opinions and addressing other group dynamic issues	[28,29]
Delphi method[α,β]	Structured, iterative forecasting method involving a panel of experts who provide anonymous responses to questionnaires with the opportunity to revise their responses when the anonymous summary of response from the prior round are revealed	[30,31]
Dyadic interview[α,β]	Method that utilizes two participants in a single interview, responding to open-ended questions asked by an interviewer to identify how a product, service or opportunity is perceived	[32,33]
Focus group[α,β]	Method that utilizes a group of interacting individuals that provide information about a specific issue to identify how a product, service or opportunity is perceived	[34,35]
In depth - individual interview[α,β]	Interview technique that allows for an intensive discussion with one interviewee to explore their perspectives on a particular topic or theme, to gain a deeper understanding of this particular topic or theme. Often only a limited amount of questions or themes are prepared by the interviewer, while the rest of the questions are based on the response of the interviewee	[36,37]
Nominal group technique[β]	Method that utilizes a group process that involves making decisions by vote and ranking responses given by members of the group	[38,39]
Public meetings[β]	Method to gain public opinions on particular issues by allowing general members of the public to attend and voice their responses	[40,41]
(Semi-)structured individual interview[α,β]	Interview technique that allows new ideas to be brought up during the interview as a result of what the interviewee says in a semi-structured setting, whereas in the structured setting the interviewer strictly sticks to an interview guide and does not ask questions based on the response of the interviewee	[42,43]
Elicitation methods		
Adaptive conjoint analysis[α]	Method similar to regular conjoint analysis, but with adaptive conjoint choice tasks are based on the earlier choices made within the survey, in theory allowing the survey to focus attention on those attributes or levels of those attributes that have the most influence on the choices of that individual. Unlike discrete choice experiments this method is founded in the theory of conjoint measurement (CM), which is more focused on the behavior of number systems instead of the behavior of human preferences.	[44,45,81]

Allocation of points$^{\beta}$	Method that involves asking respondents to rate their conditions on scales, while knowing the weights which they attach to different criteria, indicating the relative importance of particular areas of their lives	[46,47]
Analytic hierarchy process$^{\alpha,\beta}$	Method in which responders assess the relative importance of pairs of attributes (treatment endpoints, properties, criteria, items, objects, etc.) towards achieving a goal, where these responses are used to compute a weight for each attribute	[20,48]
Best-worst scaling (Types 1, 2, 3)$^{\alpha,\beta}$	Involves respondents responding to surveys that include lists of attributes or profiles and being asked to indicate the best (or most appealing/important) and the worst (or least appealing/important) of them. This method consists of three types: In type 1 a set of attributes is showed that may not reflect the characteristics of any particular treatment, of which the respondent picks the best and worst. Type 2 involves a situation in which the attributes collectively characterize a particular profile and the respondent chooses the best and worst. In type 3 three or more profiles are showed and the respondent selects the best and worst profiles	[20,49,50]
Constant sum scaling$^{\gamma}$	Constant sum scaling consists of a comparative scale where respondents are asked to allocate a fixed amount (or constant sum) of points, dollars, or anything among a set of objects according to a criterion	[51,52]
Contingent valuation$^{\alpha,\beta}$	Method to determine the willingness to pay (WTP), where individuals are presented with a choice between not having the commodity valued and having the commodity but forgoing a certain amount of money. The money being that they are willing to forgo to have the commodity is their WTP for that commodity. WTP can be calculated directly using a threshold or indirectly using a discrete choice experiment for example	[53,54]
Control preference scale$^{\alpha}$	The control preferences scale (CPS) is a method to determine the degree of control a patient wants regarding medical treatment. The preference orders are analyzed using unfolding theory to determine the distribution of preferences in different populations and the effect of covariates on consumer preferences	[55,56,57]
Discrete choice experiment$^{\alpha,\beta}$	Method that utilizes an attribute-based measure of benefit, during which individuals are offered a series of hypothetical choice situations (i.e., choice sets), from which they are asked to choose between two or more profiles. There are numerous variants of discrete choice experiments. In contrast to conjoint analysis, this method relies on a theory of the behavior of human preferences (for example random utility theory (RUM))	[58,59,60, 81]
Measure of value$^{\beta}$	Method used to identify the optimal bundle of services to be provided given resource constraints. Individuals are asked to allocate a fixed amount of resources between different services. These allocations are analyzed to identify the trade-offs individuals make	[61]
Outcome prioritization tool$^{\alpha}$	Instrument that allows participants to prioritize outcomes making use of a specific tool according to the “trade-off” principle, implying that they are willing to compromise on the less important outcomes	[62]
Person-trade off$^{\alpha,\beta}$	An extension of the time trade-off. With person trade-off an individual evaluates the health effects of interventions using persons (instead of time) as the equilibrating mechanism.	[63,64]
(Probabilistic)Threshold technique$^{\alpha,\beta}$	Method that determines the maximal change in one attribute respondents are willing to accept to achieve a given change in another attribute	[20,65]
Q-methodology$^{\gamma}$	Method which uses a specially designed response grid to present respondents with a set of statements and asking them to order, usually based on the extent to which they agree with them	[66,67]
Qualitative discriminant process$^{\beta}$	Method that involves a scoring and ranking process based on decision analysis technique, involving the definition of options in terms of qualitative categories, then deriving a numeric point estimate, and finally solving a maximization problem with given constraints	[68]

Repertory grid method[α]	Method used for eliciting personal constructs, i.e. what people think about a given topic. To identify preferences overlapping and rating techniques are used	[69,70]
Self-explicated conjoint[γ]	Method that asks explicitly about the preference for each attribute rather than the preference of several	[71]
Standard gamble[α,β]	Method in which respondents are asked to choose between a certain outcome and a gamble which may result in either a better outcome with a probability p, or a worse outcome than the original with a probability 1-p	[72,73]
Starting known efficacy[α]	Method similar to (probabilistic) threshold techniques, but with a specific known starting point. This method is specifically used within the context of the medical product lifecycle	[74]
Swing weighting[β]	Method for setting the weights in which a decision-relevant range is specified for each attribute, and the impact of "swinging" the attribute through that entire range of values is assigned a weight relative to the impact of swinging the attribute with the largest weight	[19,20]
Test trade-off[γ]	Method that can be regarded as an extension of the time trade-off which is specifically used to evaluate a new biomarker by using risks (instead of time) as the equilibrating mechanism	[75,76]
Time trade-off[α,β]	Method that presents individuals with a choice between living for a period in a specified, but less than perfect, state versus having a healthier life for a period of time, where time is varied until the respondent is in-different between the alternatives	[20,77,78]
Visual analogue scale[α,β]	A self-reporting instrument consisting of a line of predetermined length that separates extreme boundaries of the phenomenon being measured	[79,80]

[α] Identified in systematic review (19 methods)
[β] Identified through analysis of previous preference method reviews (23 methods)
[γ] Identified with expert consultations (4 methods)
In total 32 unique methods were identified

2.3 Taxonomy of preference methods

There are many ways to group preference methods. In this study, we grouped the identified methods according to their manner of data collection and the similarities in their method of analysis. This grouping was not intended to be a formal lexicon, but primarily served as a taxonomy to organize results and to develop a compendium of preference exploration and elicitation methods. Preference exploration methods can be grouped according to the number of participants the method utilizes in one session (Figure 1). (Semi-)Structured individual interviews, in-depth interviews, and complaints procedures use interviews with one participant (n = 1) in a single setting or session. Delphi method, focus groups, dyadic interviews, public meetings, nominal group technique, and citizen juries typically direct questions to more than one participant (n > 1) in a single setting. Concept mapping can employ either individual or group settings for data collection (n ≥ 1).

Preference elicitation methods can be grouped into four distinct groups (Figure 2), with methods from left to right being able to answer a smaller subset of research questions (a DCE is for example able to provide both willingness-to-pay (WTP) information and probability scores while contingent valuation provides WTP information only). Firstly, discrete choice-based methods typically examine the importance of trade-offs between attributes and their alternatives through a series of choice sets that present (hypothetical) alternatives. Secondly, ranking (or related) methods were classified based on the use of ranking exercises to capture the order of alternatives or attributes within a presented set. Thirdly, indifference techniques are methods that vary the value of one attribute in one of the alternatives until the participant is indifferent, or has no preference, between alternatives. Lastly, rating (or related) methods, are methods based on their utilization of comparative rating approaches, often allowing participants to express the strength of their preferences along a labeled scale.

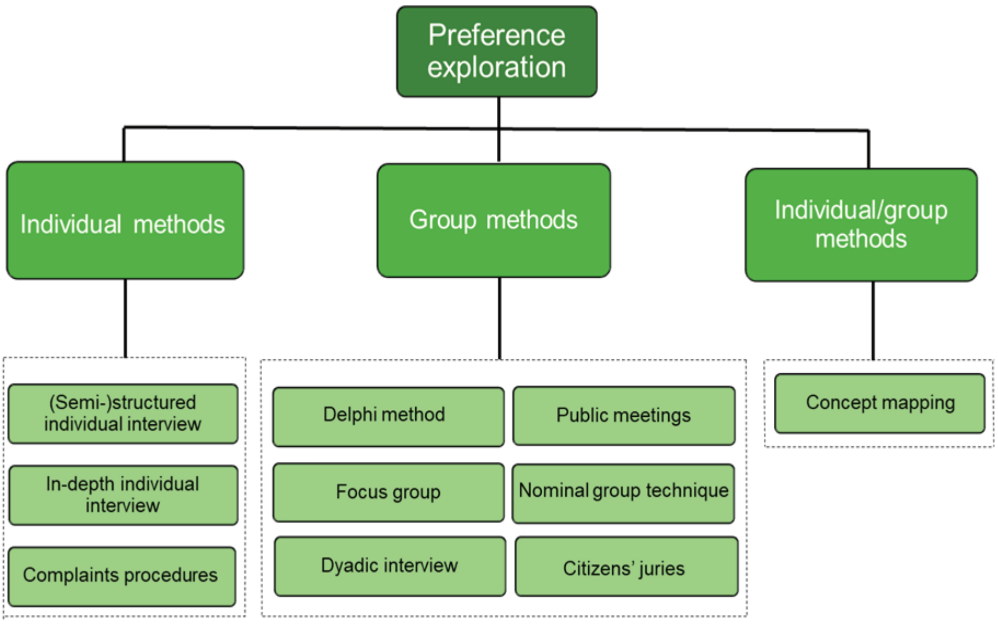


Figure 1 Grouping of preference exploration (qualitative) methods into three groups: individual, group and individual/group methods

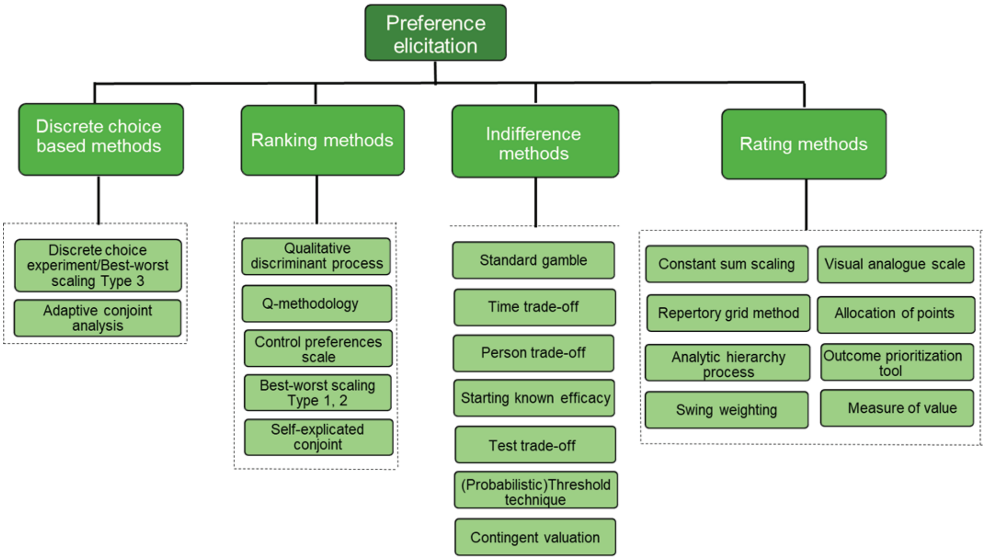


Figure 2 Grouping of preference elicitation (quantitative) methods into four groups: discrete choice based, ranking, indifference and rating methods

2.4 Trends in the use of preference methods

With the systematic literature review, spanning 37 years of literature, we observed an overall upwards trend in the number of MPLC patient preferences studies per year. The mean number of preference studies increased from 1.1 per year to 6.5 per year to 20.3 per year. This is for the periods 1980-2000, 2001-2010, and 2011-2016, respectively (Figure available in Appendix C of the supplementary material). We also observed that our included papers originated from all over the world, covering five different continents (Table 2). The majority (73%) of papers were from North America (n=90) and Europe (n=62).

Analyzing the separate use of preference exploration and elicitation methods over time, we observed a trend of preference exploration methods being used more frequently in recent years. We did not consider the period 1980 until 2005 since this period only included a few data points to compute representative percentages. For the period 2002 until 2006, 33.3% of the papers used a preference exploration method to gain insights into patient preferences (computed as the frequency of both an exploration or elicitation method in each individual paper). This increased to 48.8% in the period 2007-2011 and to 45.8% for 2012-2016. Amongst preference exploration methods, the proportion of studies that used focus groups increased from 23.0% in the period 2002-2006, to 35.0% in the period 2012-2016. The proportion of (semi-)structured individual interviews remained more or less constant with 55.0% in the period 2002-2006 and 52.0% in the period 2012-2016, while in-depth individual interviews decreased from 23.0% in 2002-2006 to 8.0% in 2012-2016. Over time, we also observed more diversity within the group of preference exploration methods. The delphi method and dyadic interviews began appearing in 2007.

Amongst preference elicitation methods, we observed that the number of papers that made use of a discrete choice experiment increased from 38.0% in 2002-2006 to 58.0% in 2012-2016. Papers that included a visual analogue scale decreased from 16.0% to 3.0%, and

Table 2 Background information of identified patient preference methods in the systematic review focusing on the medical product lifecycle

Method	Frequency		Continents of origin	Study numbers
n=19	n=208*	(%)	Continents (frequency)*	n=208
Exploration methods				
Delphi method	3	(1.4)	Asia (2), North America (1)	24, 107, 308
Dyadic interview	1	(0.5)	Africa (1)	269
Focus group	29	(13.9)	Africa (1), Asia (2), Australia/Oceania (3), Europe (15), North America (8)	2, 14, 17, 18, 43, 45, 71, 72, 84, 97, 109, 116, 119, 121, 211, 220, 222, 236, 253, 269, 282, 283, 286, 290, 294, 300, 308, 313, 317
In depth - individual interview	9	(4.3)	Asia (1), Australia/Oceania (1), Europe (3), North America (4)	32, 41, 108, 147, 173, 191, 193, 211, 316
(Semi-)structured individual interview	47	(22.6)	Africa (2), Asia (6), Australia/Oceania (6), Europe (18), North America (15)	2, 9, 17, 18, 21, 30, 41, 43, 57, 58, 65, 67, 87, 94, 100, 101, 120, 129, 141, 153, 162, 164, 184, 193, 198, 205, 211, 215, 217, 222, 226, 229, 230, 232, 239, 267, 268, 269, 272, 280, 284, 285, 286, 302, 306, 310, 323
Elicitation methods				
Adaptive conjoint analysis	3	(1.4)	North America (3)	88, 89, 243
Analytic hierarchy process	1	(0.5)	Europe (1)	221
Best-worst scaling (Types 1, 2, 3)	4	(1.9)	Asia (1), Australia/Oceania (1), North America (2)	133, 180, 189, 300
Contingent valuation	11	(5.3)	Asia (2), Australia/Oceania (1), North America (2)	29, 35, 144, 148, 155, 166, 167, 180, 199, 244, 298
Control preference scale	3	(1.4)	Asia (1), North America (2)	147, 175, 316
Discrete choice experiment	57	(27.4)	Africa (1), Asia (7), Australia/Oceania (6), Europe (15), North America (28)	19, 25, 26, 34, 42, 48, 57, 66, 73, 79, 80, 90, 100, 101, 109, 114, 117, 119, 122, 133, 134, 154, 155, 160, 161, 163, 166, 179, 180, 184, 192, 194, 200, 212, 213, 215, 218, 219, 222, 227, 229, 234, 238, 239, 243, 246, 247, 249, 257, 264, 266, 272, 281, 309, 311, 312, 313
Outcome prioritization tool	1	(0.5)	Europe (1)	304
Person-trade off	1	(0.5)	Europe (1)	274
Repertory grid method	1	(0.5)	Europe (1)	255
Standard gamble	11	(5.3)	Asia (1), Australia/Oceania (1), Europe (2), North America (7)	34, 42, 155, 180, 195, 200, 209, 219, 237, 277, 312
Starting known efficacy	1	(0.5)	North America (1)	201
(Probabilistic)Threshold Technique	2	(1.0)	North America (2)	42, 172
Time trade-off	11	(5.3)	Australia/Oceania (1), Europe (2), North America (8)	33, 34, 78, 155, 180, 200, 209, 219, 237, 277, 318
Visual analogue scale	12	(5.8)	Asia (2), Europe (3), North America (7)	93, 115, 168, 171, 178, 195, 208, 223, 278, 281, 287, 314

* Included countries per continent: Africa = Kenya, South Africa; Asia = China, Iran, Japan, Malaysia, Singapore, South Korea, Taiwan, Thailand, Turkey; Australia/Oceania = Australia; Europe = France, Germany, Great Britain, Hungary, Netherlands, Norway, Spain; North America = Canada, United States of America

contingent valuation showed a similar trend (17.0% to 9.0%). Standard gamble and time trade-off showed an upwards trend, from 5.0% and 4.0% in 2002-2006 to 9.0% and 6.0% in 2012-2016, respectively. Overall, we observed that over time, a more diverse group of preference elicitation methods was used.

2.5 Comparison of sources

The results of this study were partly in line with the results found by Ryan *et al.* (2001) and the MDIC (2015) [19,20]. Fifty-six percent (13 out of 23) methods reported by Ryan *et al.* [19] and/or the MDIC [20] were identified in our systematic literature review. The differences are due to (1) the search in this study focused specifically on methods to obtain patient preferences for drugs and medical devices, while Ryan *et al.* [19] focused on public views on the provision of healthcare, (2) MDIC [20] excluded preference exploration methods and (3) the MDIC [20] effort did not use a systematic approach for identifying methods. The taxonomy of preference methods proposed in this study is also in line with results from Mt-Isa *et al.* [21], Zhang *et al.* [22] and Gonzalez *et al.* [23], in which elicitation methods were grouped in rating, ranking and trade-off (which included choice-based techniques) techniques, although many other ways to group these methods are possible.

Results from our study's systematic literature review (19 preference methods identified) showed that most reviewed papers used focus groups, (semi-)structured individual interviews, discrete choice experiments, or the visual analogue scale to gain insights into patient preferences. Most of these studies were conducted in North-America or Europe. We also showed that the mean number of patient preference studies for drugs and medical devices increased over time. Furthermore, this study showed that for both preference exploration and elicitation methods, a more diverse mix of methods (both exploration and elicitation methods) was used over time to explore or elicit preferences.

2.6 Concluding remarks

In this study, we developed an up-to-date compendium and taxonomy of preference exploration and elicitation methods in the context of the MPLC. The systematic review (19 methods), analysis of prior conducted preference method reviews (23 methods), and expert consultations (4 methods) contributed to this compendium. In total, 32 unique methods were identified. Preference exploration methods were grouped into three main groups, while the preference exploration methods were grouped in four main groups. Since choosing which method to use will depend on the MPLC phase and what the measured preferences are being used for, future research might focus on determining which methods are most appropriate to explore or elicit patient preferences, and under what circumstances, throughout the different phases in the MPLC. In addition, it may be of interest for future research to focus on the specific combinations of preference exploration and elicitation methods used in mixed-method studies, and the reasoning behind such study designs.

References

[1] Hoos, A., Anderson, J., Boutin, M., Dewulf, L., Geissler, J., Johnston, G., et al. Partnering With Patients in the Development and Lifecycle of Medicines: A Call for Action. Ther Innov Regul Sci 2015, 49(6), 929-939. doi: 10.1177/2168479015580384.

[2] Anderson, M. & McCleary, K., On the path to a science of patient input. Science Translational Medicine 2016, 8(336): p. 336ps11. doi: 10.1126/scitranslmed.aaf6730.

[3] Smith, M.Y., Hammad, T.A., Metcalf, M., Levitan, B., Noel, R., Wolka A.N., Michaels, D.L. & Johnson, F.R. Patient Engagement at a Tipping Point—The Need for Cultural Change Across Patient, Sponsor, and Regulator Stakeholders:Insights From the DIA Conference, "Patient Engagement in Benefit Risk Assessment Throughout the Life Cycle of Medical Products". Therapeutic Innovation & Regulatory Science 2016, 50(5), 546-553.

[4] Bekker-Grob, de, E. W., Berlin, C., Levitan, B., Raza, K., Christoforidi, K., Cleemput, I., et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. The Patient - Patient-Centered Outcomes Research 2017, 10(3), 263-266. doi: 10.1007/s40271-017-0222-3.

[5] Anderson, R. M., & Funnell, M. M. Patient empowerment: reflections on the challenge of fostering the adoption of a new paradigm. Patient Education and Counseling 2005, 57(2), 153-157. doi: https://doi.org/10.1016/j.pec.2004.05.008.

[6] van Overbeeke, E., Whichello, C., Janssens, R., Veldwijk, J., Cleemput, I., Simoens, S., et al. Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. Drug Discovery Today 2018. https://doi.org/10.1016/j.drudis.2018.09.015

[7] FDA. Guidance for Industry, Food and Drug Administration Staff and Other Stakeholders. Patient Preference Information –Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. Public report 2016a.

[8] FDA. Public Workshop On Patient-Focused Drug Development: Guidance 1 Collecting Comprehensive And Representative Input. Public report 2017a.

[9] FDA. The Voice of the Patient: A Series of Reports from FDA's Patient-Focused Drug Development Initiative. Public report 2017b.

[10] FDA. PDUFA Reauthorization Performance Goals And Procedures Fiscal Years 2018 Through 2022. Public report 2017c.

[11] Kievit W., Tummers M., van Hoorn R., Booth A., Mozygemba K., Refolo P., et al. Taking patient heterogeneity and preferences into account in health technology assessments. Int. J. Technol. Assess. Health Care, 33 (2017), pp. 562-569.

[12] Abelson J., Wagner F., DeJean D., Boesveld S., Gauvin F.P., Bean S., et al. Public and patient involvement in health technology assessment: a framework for action. Int. J. Technol. Assess. Health Care, 32 (2016), pp. 256-264.

[13] Lowe M.M., Blaser D.A., Cone L., Arcona S., Ko J., Sasane R., et al. Increasing patient involvement in drug development Value Health, 19 (2016), pp. 869-878.

[14] Stewart K.D., Johnston J.A., Matza L.S., Curtis S.E., Havel H.A., Sweetana S.A., et al. .Preference for pharmaceutical formulation and treatment process attributes Patient Prefer. Adherence, 10 (2016), pp. 1385-1399.

[15] Minion L.E., Coleman R.L., Alvarez R.D., Herzog T.J. Endpoints in clinical trials: what do patients consider important? A survey of the Ovarian Cancer National Alliance Gynecol. Oncol., 140 (2016), pp. 193-198.

[16] Bloom D., Beetsch J., Harker M., Hesterlee S., Moreira P., Patrick-Lake B., et al. The rules of engagement: CTTI recommendations for successful collaborations between sponsors and patient groups around clinical trials Ther. Innov. Regul. Sci., 52 (2018), pp. 206-213.

[17] Beshears, J., Choi JJ., Laibson D., & Madrian BC. How Are Preferences Revealed? Journal of Public Economics 2008, 92(8-9):1787-1794. doi:10.1016/j.jpubeco.2008.04.010.

[18] Lambooij, M. S., Harmsen, I. A., Veldwijk, J., de Melker, H., Mollema, L., van Weert, Y. W., & de Wit, G. A. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Medical Research Methodology 2015, 15, 19. doi:http://doi.org/10.1186/s12874-015-0010-5.

[19] Ryan, M., Scott, D. A., Reeves, C., Bate, A., van Teijlingen, E. R., Russell, E., et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 2001, 5(5).

[20] Medical Device Innovation Consortium (MDIC). Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public report 2015.

[21] Mt-Isa, S., Hughes, D., Waddingham, E., Chan, E., Hallgreen, C., Downey G., Hockley, K., Juhaeri, J., Lieftucht, Alfons., Metcalf, M., Noel, R., Philips, L., Ashby, D. & Micaleff, A. IMI-PROTECT Benefit-Risk Group RECOMMENDATIONS REPORT: Recommendations for the methodology and visualisation techniques to be used in the assessment of benefit and risk of medicines. IMI PROTECT Recommendations Report 2013.

[22] Zhang, W., Ji, L., Chen, Y., Tang, K., Wang, H., Zhu, R., Jia, W., Cao, Z. & Liu, Q. When drug discovery meets web search: Learning to Rank for ligand-based virtual screening. Journal of Cheminformatics 2015, 7.

[23] Gonzalez, J.M., Johnson, F.R., McAteer, H., Poser, J. & Mughal, F. Comparing preferences for outcomes of psoriasis treatments among patients and dermatologists in the U.K.: results from a discrete-choice experiment. British Journal of Dermatology 2017, 176(3), 777-785.

[24] Mitton, C., Smith, N., Peacock, S., Evoy, B., & Abelson, J. Public participation in health care priority setting: A scoping review. Health Policy 2009, 91(3), 219-228. doi:10.1016/j.healthpol.2009.01.005.

[25] Street, J., Duszynski, K., Krawczyk, S., & Braunack-Mayer, A. The use of citizens' juries in health policy decision-making: A systematic review. Social Science & Medicine 2014, 109, 1-9. doi:http://dx.doi.org/10.1016/j.socscimed.2014.03.005

[26] Bourne, T., Wynants, L., Peters, M., Van Audenhove, C., Timmerman, D., Van Calster, B., et al. The impact of complaints procedures on the welfare, health and clinical practise of 7926 doctors in the UK: a cross-sectional survey. BMJ Open 2015, 5.

[27] Wensing, M., & Grol, R.What can patients do to improve health care? Health Expect 1998, 1(1), 37-49.

[28] Burke, J. G., O'Campo, P., Peak, G. L., Gielen, A. C., McDonnell, K. A., & Trochim, W. M. K. An Introduction to Concept Mapping as a Participatory Public Health Research Method. Qualitative Health Research 2005, 15(10), 1392-1410. doi:doi:10.1177/1049732305278876.

[29] Trochim, W., & Kane, M. Concept mapping: an introduction to structured conceptualization in health care. International Journal for Quality in Health Care 2005, 17(3), 187-191. doi:10.1093/intqhc/mzi038.

[30] Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. PLOS ONE 2005, 6(6), e20476. doi:10.1371/journal.pone.0020476.

[31] Julian de, M.. The Delphi method and health research. Health Education 2003, 103(1), 7-16. doi:10.1108/09654280310459112.

[32] Eisikovits, Z., & Koren, C. Approaches to and Outcomes of Dyadic Interview Analysis. Qualitative Health Research 2010, 20(12), 1642-1655. doi:10.1177/1049732310376520

[33] Morgan, D. L., Ataie, J., Carder, P., & Hoffman, K. Introducing Dyadic Interviews as a Method for Collecting Qualitative Data. Qualitative Health Research 2013, 23(9), 1276-1284. doi:10.1177/1049732313501889.

[34] Basch, C. E. Focus Group Interview: An Underutilized Research Technique for Improving Theory and Practice in Health Education. Health Education Quarterly 1987, 14(4), 411-448. doi:10.1177/109019818701400404.

[35] Schulze, B., & Angermeyer, M. C. Subjective experiences of stigma. A focus group study of schizophrenic patients, their relatives and mental health professionals. Social Science & Medicine 2003, 56(2), 299-312. doi:http://dx.doi.org/10.1016/S0277-9536(02)00028-X.

[36] Harris, M. A., Wysocki, T., Sadler, M., Wilkinson, K., Harvey, L. M., Buckloh, L. et al. Validation of a structured interview for the assessment of diabetes self-management. Diabetes Care 2000, 23(9), 1301.

[37] Williams, J. B. A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 1988 (0003-990X (Print)).

[38] Allen, J., Dyas, J., & Jones, M. Building consensus in health care: a guide to using the nominal group technique. British Journal of Community Nursing 2004, 9(3), 110-114. doi:10.12968/bjcn.2004.9.3.12432.

[39] Gallagher, M., Hares, T. I. M., Spencer, J., Bradshaw, C., & Webb, I. A. N. The Nominal Group Technique: A Research Tool for General Practice? Family Practice 1993, 10(1), 76-81. doi:10.1093/fampra/10.1.76.

[40] Ham, C. Priority setting in health care: learning from international experience. Health Policy 1997, 42(1), 49-66. doi:http://dx.doi.org/10.1016/S0168-8510(97)00054-7.

[41] McComas, K. A. Theory and practice of public meetings. Communication Theory 2001, 11(1), 36-55. doi:10.1111/j.1468-2885.2001.tb00232.x.

[42] Barriball, L.K., & While, A. Collecting data using a semi-structured interview: a discussion paper. Journal of Advanced Nursing 1994, 19(2), 328-335. doi:10.1111/j.1365-2648.1994.tb01088.x.

[43] Whiting, L. S. Semi-structured interviews: guidance for novice researchers. Nursing Standard 2008, 22(23), 35-40. doi:10.7748/ns2008.02.22.23.35.c6420.

[44] Beusterien, K. M., Dziekan, K., Flood, E., Harding, G., & Jordan, J. C. Understanding Patient Preferences for HIV Medications Using Adaptive Conjoint Analysis: Feasibility Assessment. Value in Health 2005, 8(4), 453-461. doi:http://dx.doi.org/10.1111/j.1524-4733.2005.00036.x.

[45] Fraenkel, L., Bodardus, S., & Wittink, D. R. Understanding Patient Preferences for the Treatment of Lupus Nephritis With Adaptive Conjoint Analysis. Medical Care 2001, 39(11).

[46] Haywood, K. L., Garratt, A. M., Dziedzic, K., & Dawes, P. T. Patient centered assessment of ankylosing spondylitis-specific health related quality of life: evaluation of the Patient Generated Index. The Journal of Rheumatology 2003, 30(4), 764.

[47] Schwappach, D. L. B., & Strasmann, T. J. Quick and dirty numbers?: The reliability of a stated-preference technique for the measurement of preferences for resource allocation. Journal of Health Economics 2006, 25(3), 432-448. doi:http://dx.doi.org/10.1016/j.jhealeco.2005.08.002.

[48] Liberatore, M. J., & Nydick, R. L. The analytic hierarchy process in medical and health care decision-making: A literature review. European Journal of Operational Research 2006, 189(1), 194-207. doi:http://dx.doi.org/10.1016/j.ejor.2007.05.001.

[49] Flynn, T. N. Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. Expert Review of Pharmacoeconomics & Outcomes Research 2010, 10(3), 259-267. doi:10.1586/erp.10.29

[50] Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. Best–worst scaling: What it can do for health care research and how to do it. Journal of Health Economics 2007, 26(1), 171-189. doi:http://dx.doi.org/10.1016/j.jhealeco.2006.04.002.

[51] Mai, R., & Hoffmann, S. Taste lovers versus nutrition fact seekers: How health consciousness and self-efficacy determine the way consumers choose food products. Journal of Consumer Behaviour 2012, 11(4), 316-328. doi:10.1002/cb.1390.

[52] Skedgel, C. D., Wailoo, A. J., & Akehurst, R. L. Choosing vs. allocating: discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences. Health Expectations 2015, 18(5), 1227-1240. doi:10.1111/hex.12098.

[53] Bärnighausen, T., Liu, Y., Zhang, X., & Sauerborn, R. Willingness to pay for social health insurance among informal sector workers in Wuhan, China: a contingent valuation study. BMC Health Services Research 2007, 7(1), 114. doi:10.1186/1472-6963-7-114.

[54] Cunningham, S. J., & Hunt, N. P. Relationship between utility values and willingness to pay in patients undergoing orthognathic treatment. Community dental health 2000, 17(2), 92-96.

[55] Degner, L. F., Sloan, J. A., & Venkatesh, P. The Control Preferences Scale. Can J Nurs Res 1997.

[56] McPherson, C. J., Higginson, I. J., & Hearn, J. Effective methods of giving information in cancer: a systematic literature review of randomized controlled trials. Journal of Public Health 2001, 23(3), 227-234. doi:10.1093/pubmed/23.3.227.

[57] Henrikson, N.B., Davison, B.J., Berry, D.L. Measuring decisional control preferences in men newly diagnosed with prostate cancer. Journal of Psychosocial Oncology 2011, 29(6), 606-618.

[58] Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., 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), 403-413.

doi:http://dx.doi.org/10.1016/j.jval.2010.11.013.

[59] Soekhai, V., Bekker-Grob, E.W. de, Ellis, A.R., Vass, C.M. Discrete Choice Experiments in Health Economics: Past, Present and Future. PharmacoEconomics 2018. https://doi.org/10.1007/s40273-018-0734-2

[60] Lancsar, E., & Louviere, J. Conducting Discrete Choice Experiments to Inform Healthcare Decision-making. PharmacoEconomics 2008, 26(8), 661-677. doi:10.2165/00019053-200826080-00004.

[61] Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., & Luce, B. R. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value in Health 2003, 6(1), 9-17. doi:http://dx.doi.org/10.1046/j.1524-4733.2003.00234.x

[62] Case, S. M., O'Leary, J., Kim, N., Tinetti, M. E., & Fried, T. R. Older Adults' Recognition of Trade-Offs in Healthcare Decision-Making. J Am Geriatr Soc 2015, 63(8), 1658-1662. doi:10.1111/jgs.13534.

[63] Green, C. On the societal value of health care: what do we know about the person trade-off technique? Health Econ 2001, 10(3), 233-243. doi:10.1002/hec.581.

[64] Nord, E. The Person-trade-off Approach to Valuing Health Care Programs. Medical Decision-making 1995, 15(3), 201-208. doi:10.1177/0272989X9501500302.

[65] Kopec, J. A., Richardson, C. G., Llewellyn-Thomas, H., Klinkhoff, A., Carswell, A., & Chalmers, A. Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied. Journal of Clinical Epidemiology 2007, 60(9), 929-938. doi:http://dx.doi.org/10.1016/j.jclinepi.2007.01.001.

[66] Cross, R. M. Exploring attitudes: the case for Q methodology. Health Education Research 2005, 20(2), 206-213. doi:10.1093/her/cyg121.

[67] Exel, J. van., Graaf, G. d., & Brouwer, W. Care for a break? An investigation of informal caregivers' attitudes toward respite care using Q-methodology. Health Policy 2007, 83(2–3), 332-342. doi:http://dx.doi.org/10.1016/j.healthpol.2007.02.002.

[68] Mullen, P. M. Public involvement in health care priority setting: an overview of methods for eliciting values. Health Expect 1999, 2(4), 222-234.

[69] Davis, C. J., Fuller, R. M., Tremblay, M. C., & Berndt, D. J. Communication Challenges in Requirements Elicitation and the Use of the Repertory Grid Technique. Journal of Computer Information Systems 2006, 46(5), 78-86. doi:10.1080/08874417.2006.11645926.

[70] Rowe, G., Lambert, N., Bowling, A., Ebrahim, S., Wakeling, I., & Thomson, R. Assessing patients' preferences for treatments for angina using a modified repertory grid method. Social Science & Medicine 2005, 60(11), 2585-2595. doi:http://dx.doi.org/10.1016/j.socscimed.2004.10.019.

[71] Riquelme, H., & Rickards, T. Hybrid conjoint analysis: An estimation probe in new venture decisions. Journal of Business Venturing 1992, 7(6), 505-518. doi:http://dx.doi.org/10.1016/0883-9026(92)90022-J.

[72] Gafni, A. The standard gamble method: what is being measured and how it is interpreted. Health Services Research 1994, 29(2), 207-224.

[73] Morimoto, T., & Fukui, T. Utilities Measured by Rating Scale, Time Trade-off, and Standard Gamble: Review and Reference for Health Care Professionals. Journal of Epidemiology 2002, 12(2), 160-178. doi:10.2188/jea.12.160.

[74] Man-Son-Hing, M., Laupacis, A., O'Connor, A., Wells, G., Lemelin, J., Wood, W., & Dermer, M. Warfarin for atrial fibrillation. The patient's perspective. Arch Intern Med 1996, 156(16), 1841-1848.

[75] Baker, S. G., & Kramer, B. S. evaluating surrogate endpoints, prognostic markers, and predictive markers: Some simple themes. Clinical Trials 2014, 12(4), 299-308. doi:10.1177/1740774514557725.

[76] Baker, S. G., Van Calster, B., & Steyerberg, E. W. Evaluating a new marker for risk prediction using the test tradeoff: an update. Int J Biostat 2012, 8(1). doi:10.1515/1557-4679.1395.

[77] Arnesen, T., & Trommald, M. Are QALYs based on time trade-off comparable?--A systematic review of TTO methodologies. Health Econ 2005, 14(1), 39-53. doi:10.1002/hec.895.

[78] Brazier, J., Deverill, M., & Green, C. A Review of the Use of Health Status Measures in Economic Evaluation. Journal of Health Services Research & Policy 1999, 4(3), 174-184. doi:10.1177/135581969900400310.

[79] Holdgate, A., Asha, S., Craig, J., & Thompson, J. Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain. Emerg Med (Fremantle) 2003, 15(5-6), 441-446.

[80] Räsänen, P., Roine, E., Sintonen, H., Semberg-Konttinen, V., Ryynänen, O.-P., & Roine, R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. International Journal of Technology Assessment in Health Care 2006, 22(2), 235-241. doi:10.1017/S0266462306051051.

[81] Louviere, J., Flynn, T., Carson, R. Discrete Choice Experiments Are Not Conjoint Analysis. Journal of Choice Modelling 2010, 3(3), 57-72.

Appendices

Appendix A: Methods used in the study

A.1 Systematic literature review

The systematic literature review (Figure A) was conducted in accordance with standard guidelines [A1], searching for relevant peer-reviewed papers from 1980 until 2016 in six different scientific databases: Cochrane Library, Econlit, Embase, PubMed, PsycInfo and Scopus. We used broad search terms to minimize the chance of missing methods. The following search query was developed to search for papers in the databases Cochrane Library, Econlit, Embase, PubMed, PsycInfo and Scopus (this query is based on the PubMed database and was translated to other databases when necessary):

((patient[tiab] OR patients[tiab] OR client[tiab] OR clients[tiab] OR citizen[tiab] OR citizens[tiab] OR consumer[tiab] OR consumers[tiab] OR public[tiab] OR general population[tiab]) AND (Preference[tiab] OR Preferences[tiab] OR Acceptability[tiab] OR utility[tiab] OR utilities[tiab] OR desirability[tiab]) AND (method[tiab] OR methods[tiab] OR methodology[tiab] OR methodologies[tiab] OR technique[tiab] OR techniques[tiab] OR measurement[tiab] OR measurements[tiab] OR instrument[tiab] OR instruments[tiab] OR tool[tiab] OR tools[tiab] OR measure[tiab] OR measures[tiab] AND (elicitation[tiab] OR elicitations[tiab] OR measuring[tiab] OR determine[tiab] OR determining[tiab] OR explore[tiab]) AND (drug[tiab] OR drugs[tiab] OR pharmaceutical[tiab] OR pharmaceuticals[tiab] OR medicine[tiab] OR medicines[tiab] OR medication[tiab] OR medications[tiab] OR “medical device”[tiab] OR “medical devices”[tiab] OR “medical instrument”[tiab] OR “medical instruments”[tiab])) AND ("loattrfull text"[sb] AND ("1980/01/01"[PDAT] : "2016/12/31"[PDAT]) AND "humans"[MeSH Terms] AND English[lang]).

Studies were included if they focused on technical/methodological aspects of preference exploration or elicitation methods (including papers focused on reviews or protocols) or were related to a (technical) application study, drugs or medical devices, patient (or caregivers) preferences, and if the full text was available in English. Papers were excluded if they included instruments specifically developed for the quality-adjusted life year (QALY) framework (e.g. EQ-5D, HUI, SF-36) and no specific preference method was used to generate QALY weights, since these methods are generally not able to gain insights into patient preferences outside of the QALY framework. Papers were screened and analyzed independently by two reviewers (VS and CW) based on the aforementioned inclusion criteria. Differences were discussed between both reviewers to reach consensus, and a third reviewer (JV or EBG) was consulted when consensus could not be reached.

A total of 12,136 references were identified from the beginning of 1980 until the end of 2016 (Figure A). From these references, 4,572 papers met the inclusion criteria and remained for title and abstract screening. The title and abstract screening excluded another 4,249 papers based on the inclusion criteria. Hence, 323 full-text papers were screened for eligibility. A total of 115 papers were excluded since they did not meet the inclusion criteria, did not mention a preference method, or were focused on scale related and/or decision-making framework related techniques. The 208 included papers comprised of 5 systematic literature reviews, 18 technical and/or methodological papers, and 185 (technical) application studies.

A.2 Analysis of prior reviews

Prior reviews by Ryan *et al.* [A2] and from the MDIC [A3] were analyzed for additional preference methods. These reviews were analyzed independently by two reviewers (VS and CW), and a third (JV or EBG) reviewer was consulted when there were differences and consensus could not be reached. These activities served as a cross-validation for our results from the systematic literature review.

A.3 Expert consultation rounds

To account for publication lag and to confirm our results, international experts (n=24) in the field of health preferences and/or medical decision-making were consulted. In this final step, experts were consulted via e-mail, phone interviews or in person (e.g. during the 7th Meeting of the International Academy of Health Preference Research (IAHPR)) to critically assess our list of identified methods, and the grouping of the identified methods, until data saturation was reached. The experts were from three different continents and had various, specialized areas of expertise in health preferences and/or medical decision-making.

A.4 Strengths and weaknesses

The current study has several strengths. First, the systematic literature review, followed by an analysis of prior reviews and expert consultations, resulted in an updated compendium of methods (identifying 9 methods not reported in previously-conducted reviews) to gain insights into patient preferences throughout the MPLC. Using previously-conducted reviews as validation, and consulting health preference and medical-decision-making experts to confirm our results, is important to build a well-founded and accepted basis across the field of health preference research. Secondly, the systematic literature review was conducted independently by two researchers, which had a positive effect on reaching a reliable overview of preference exploration and elicitation methods that were applied to drugs and medical devices. Thirdly, the identification of both preference exploration and elicitation methods as presented in this paper matches the trend in the literature, such as Vass *et al.* [A4] and Ikenwilo *et al.* [A5] describing a more frequent use of mixed-method approaches to gain more insights into preferences.

A potential weakness of our study design is that the systematic literature review might have missed possible/promising methods to gain insights into patient preferences. To produce a time-efficient and precise review, we limited the synonyms for a variety of search terms. However, the impact is likely to be limited given the validation with previously-conducted reviews and additional review by international experts with significant experience in health preference research. Another weakness is that some methods in this study might be considered specific variants of other preference methods (e.g. constant sum scaling, measure of value, outcome prioritization tool, and starting known efficacy) and that there may be no consensus across different scientific fields if the method can indeed be considered a "preference method" itself (e.g. control preferences scale, test-trade-off and visual analogue scale). However, we decided to include these methods in our overview, since the aim of our study was to develop a current compendium of both exploration and elicitation preference methods, without excluding methods that were within the scope of our study *a priori*.

References

[A1] Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2014 edition.

[A2] Ryan, M., Scott, D. A., Reeves, C., Bate, A., van Teijlingen, E. R., Russell, E., et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 2001, 5(5).

[A3] Medical Device Innovation Consortium (MDIC). Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public report 2015.

[A4] Vass, C.M., Rigby, D. & Payne, K. Investigating the Heterogeneity in Women's Preferences for Breast Screening: Does the Communication of Risk Matter? Value in Health 2017, Epub article.

[A5] Ikenwilo, D., Heidenreich, S., Ryan, M., Mankowski, C., Nazir, J. & Watson, V. The Best of Both Worlds: An Example Mixed Methods Approach to Understand Men's Preferences for the Treatment of Lower Urinary Tract Symptoms. The Patient: Patient-Centered Outcomes Research 2018, 11(1), 55-67.

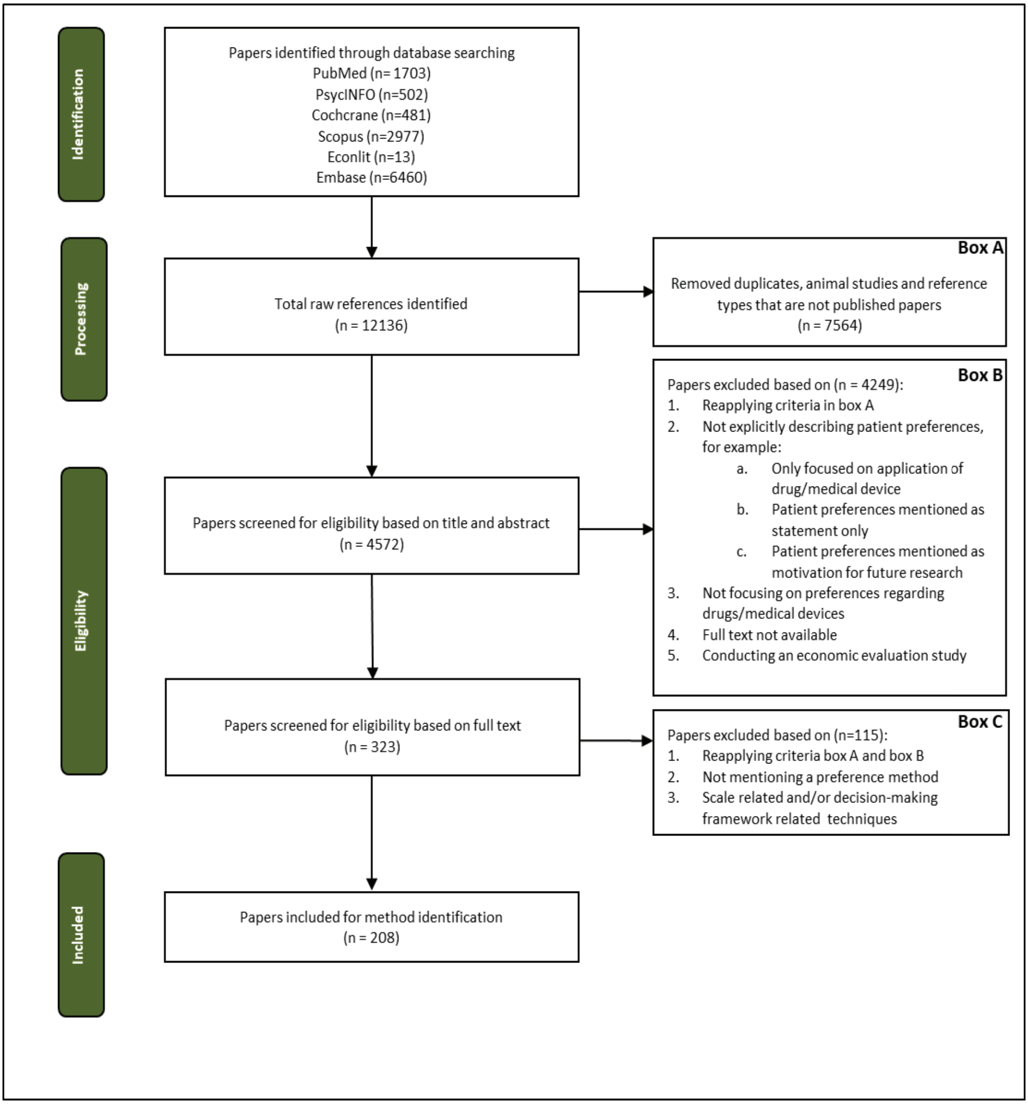


Figure A Flow diagram of systematic literature review inclusion and exclusion process

Chapter 2: Appendix B

Appendix B: Overview of analyzed full text papers in systematic literature review (n=323)

Study ID in Review	Authors	Title	Journal	Year
1*	Abraham NS, Naik AD, Street RL, et al	Complex antithrombotic therapy: Determinants of patient preference and impact on medication adherence	Patient Preference and Adherence	2015
2	Ackerman IN, Jordan JE, Van Doornum S, et al	Understanding the information needs of women with rheumatoid arthritis concerning pregnancy, post-natal care and early parenting: A mixed-methods study	BMC Musculoskelet Disord	2015
3	Ahmad Y, Nijjer S, Cook CM, et al	A new method of applying randomised control study data to the individual patient: A novel quantitative patient-centred approach to interpreting composite end points	Int J Cardiol	2015
4*	Aikens JE, Nease DE, Jr., Nau DP, et al	"Adherence to maintenance-phase antidepressant medication as a function of patient beliefs about medication": Correction	Annals of Family Medicine	2005
5*	Aikens JE, Nease DE, Jr. , Nau DP, et al	Adherence to maintenance-phase antidepressant medication as a function of patient beliefs about medication	Ann Fam Med	2005
6	Alam M, Arifeen S	A qualitative research to understand public perception of personalised cancer medicine (PCM) and its application in treatment of advanced lung cancer	Asia-Pacific Journal of Clinical Oncology	2015
7*	Aleem IS, Jalal H, Aleem IS, et al	Clinical decision analysis: Incorporating the evidence with patient preferences	Patient Preference and Adherence	2009
8*	Al-Omari B, Frisher M, Croft P, et al	Using adaptive choice based conjoint (ACBC) analysis to study patients' preferences regarding pharmaceutical treatment for osteoarthritis (OA)	Annals of the Rheumatic Diseases	2013
9	Alonso-Coello P, Montori VM, Sola I, et al	Values and preferences in oral anticoagulation in patients with atrial fibrillation, physicians' and patients' perspectives: protocol for a two-phase study	BMC Health Serv Res	2008
10	Altice FL, Mostashari F, Friedland GH	Trust and the acceptance of and adherence to antiretroviral therapy	J Acquir Immune Defic Syndr	2001
11	Ambrose PG, Hammel JP, Bhavnani SM, et al	Frequentist and Bayesian pharmacometric-based approaches to facilitate critically needed new antibiotic development: overcoming lies, damn lies, and statistics	Antimicrob Agents Chemother	2012
12	Anderson P	Patient preference for and satisfaction with inhaler devices	European Respiratory Review	2005

13*	Andrade JG, Krahn AD, Skanes AC, et al	Values and Preferences of Physicians and Patients With Nonvalvular Atrial Fibrillation Who Receive Oral Anticoagulation Therapy for Stroke Prevention	Canadian Journal of Cardiology	2016
14	Anink J, Otten MH, Gorter SL, et al	Treatment choices of paediatric rheumatologists for juvenile idiopathic arthritis: etanercept or adalimumab?	Rheumatology (Oxford)	2013
15*	Anquinet L, Raus K, Sterckx S, et al	Comparing continuous sedation until death and euthanasia: Professional caregivers' attitudes and experiences. A focus group study in flanders, belgium	Palliative Medicine	2012
16	Arendts G, Howard K, Rose JM	Allocation decisions and patient preferences in emergency medicine	Emerg Med J	2011
17	Arkell P, Ryan S, Brownfield A, et al	Patient experiences, attitudes and expectations towards receiving information about anti-TNF: A qualitative study	Rheumatology (United Kingdom)	2012
18	Arkinson J, Holbrook A, Wiercioch W	Public perceptions of physician-pharmaceutical industry interactions: A systematic review	Healthcare Policy	2010
19	Ashcroft DM, Seston E, Griffiths CE	Trade-offs between the benefits and risks of drug treatment for psoriasis: a discrete choice experiment with U.K. Dermatologists	Br J Dermatol	2006
20*	Ashton H, Nodiyal A, Green D, et al	Acupuncture or counselling: Outcomes and predictors of treatment choice in a non-statutory addiction service	Journal of Substance Use	2009
21	Astin F, Closs SJ, McLenachan J, et al	The information needs of patients treated with primary angioplasty for heart attack: an exploratory study	Patient Educ Couns	2008
22*	Athar MW, Mativo C, Landis R, et al	Communication of laboratory data and diagnostic test results to hospitalized patients: A study of preferences and recall	Patient Preference and Adherence	2016
23	Attipoe L, Ciurtin C	Does health satisfaction and perception of treatment efficacy in patients with inflammatory arthritis correlate with disability and global health state?	Annals of the Rheumatic Diseases	2015
24	Bader P, McDonald P, Selby P	An algorithm for tailoring pharmacotherapy for smoking cessation: results from a Delphi panel of international experts	Tob Control	2009

25	Baji P, Gulácsi L, Golovics PA, et al	Perceived Risks Contra Benefits of Using Biosimilar Drugs in Ulcerative Colitis: Discrete Choice Experiment among Gastroenterologists	Value in Health Regional Issues	2016
26	Baji P, Gulacsi L, Lovasz BD, et al	Treatment preferences of originator versus biosimilar drugs in Crohn's disease: discrete choice experiment among gastroenterologists	Scand J Gastroenterol	2016
27*	Bakirtas A, Kutlu A, Baccioglu A, et al	Physicians' preference for controller medication in mild persistent asthma	European Journal of Allergy and Clinical Immunology	2016
28	Bardet JD, Vo TH, Bosson JL, et al	Patients', physicians' and pharmacists' viewpoints on the implementation of medication reviews in the French primary care	International Journal of Clinical Pharmacy	2015
29	Barron AC, Lee TL, Taylor J, et al	Feasibility and construct validity of the parent willingness-to-pay technique for children with juvenile idiopathic arthritis	Arthritis Rheum	2004
30	Belcher VN, Fried TR, Agostini JV, et al	Views of older adults on patient participation in medication-related decision-making	J Gen Intern Med	2006
31*	Bell CM, Chapman RH, Stone PW, et al	An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses	Med Decis Making	2001
32	Benson J, Britten N	What effects do patients feel from their antihypertensive tablets and how do they react to them? Qualitative analysis of interviews with patients	Fam Pract	2006
33	Beukelman T, Guevara JP, Albert DA	Optimal treatment of knee monarthritis in juvenile idiopathic arthritis: a decision analysis	Arthritis Rheum	2008
34	Bewtra M, Johnson FR	Assessing patient preferences for treatment options and process of care in inflammatory bowel disease: a critical review of quantitative data	Patient	2013
35	Beyhun NE, Kosan Z, Aras A, et al	Willingness to receive the influenza A(H1N1) vaccine and its determinants among university students during the 2009 outbreak in Turkey	Eurasian Journal of Medicine	2014

36*	Bhargava JS, Patel B, Foss AJ, et al	Views of glaucoma patients on aspects of their treatment: an exploration of patient preference by conjoint analysis	Invest Ophthalmol Vis Sci	2006
37*	Bime C, Wei C, Nguyen J, et al	Asthma symptom utility index: Reliability, validity, responsiveness and the minimally important clinical difference in adult asthma patients	American Journal of Respiratory and Critical Care Medicine	2012
38*	Bislew HD, Sorensen TD	Use of focus groups as a tool to enhance a pharmaceutical care practice	Journal of the American Pharmacists Association : JAPhA	2003
39*	Blank T, Graves K, Sepucha K, et al	Understanding treatment decision-making: Contexts, commonalities, complexities, and challenges	Annals of Behavioral Medicine	2006
40*	Bonistalli L, Bardelli F, Constantini M, et al	Adjuvant chemotherapy in patients with resectable stage III colon cancer: Lifetime cost-effectiveness and cost-utility analysis	Cancer Journal	1998
41	Borgsteede SD, Karapinar-Carkit F, Hoffmann E, et al	Information needs about medication according to patients discharged from a general hospital	Patient Educ Couns	2011
42	Brett Hauber A, Fairchild AO, Reed Johnson F	Quantifying Benefit-Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature	Applied Health Economics and Health Policy	2013
43	Brockbank S, Miller N, Owen S, et al	Pretreatment information on dysphagia: Exploring the views of head and neck cancer patients	Journal of Pain and Symptom Management	2015
44	Broekhuizen H, Ijzerman MJ, Hauber AB, et al	Weighing Clinical Evidence Using Patient Preferences: An Application of Probabilistic Multi-Criteria Decision Analysis	PharmacoEconomics	2016
45	Bryson SP	Patient-centred, administration friendly medicines for children - an evaluation of children's preferences and how they impact medication adherence	Int J Pharm	2014
46	Buttery A, Lowton K, Glaser K, et al	Older heart failure patients' preferences for cardiac rehabilitation service models	European Heart Journal	2011
47*	Caffey M, Maria S, Ireland M, et al	Antiemetic management preferences for Australasian paramedic providers	Australasian Journal of Paramedicine	2016

48	Cai QF, Wan F, Dong XY, et al	Fertility clinicians and infertile patients in China have different preferences in fertility care	Hum Reprod	2014
49*	Carlesso LC, MacDermid JC, Gross AR, et al	Treatment preferences amongst physical therapists and chiropractors for the management of neck pain: Results of an international survey	Chiropractic and Manual Therapies	2014
50	Casarett D, Van Ness PH, O'Leary JR, et al	Are patient preferences for life-sustaining treatment really a barrier to hospice enrollment for older adults with serious illness?	J Am Geriatr Soc	2006
51*	Castro JG, Jones DL, Weiss SM	STD patients' preferences for HIV prevention strategies	HIV/AIDS - Research and Palliative Care	2014
52	Celio J, Maeder S, Halabi G, et al	Chronic dialysis, medication adherence and beliefs about medicines: A comparison between patients born in Switzerland and migrant patients (diana study)	International Journal of Clinical Pharmacy	2016
53*	Charles C, Gafni A	The vexing problem of defining the meaning, role and measurement of values in treatment decision-making	J Comp Eff Res	2014
54*	Chater AM, Parham R, Riley S, et al	Profiling patient attitudes to phosphate binding medication: a route to personalising treatment and adherence support	Psychol Health	2014
55	Chatterjee A, DePriest K, Blair R, et al	Results of a survey of blood pressure monitoring by intensivists in critically ill patients: a preliminary study	Crit Care Med	2010
56*	Chaudhry IB, Rahman R, Minhas HM, et al	Which antidepressant would psychiatrists and nurses from a developing country choose for themselves?	nt J Psychiatry Clin Pract	2011
57	Chen LC, Cheng LJ, Zhang Y, et al	Acupuncture or low frequency infrared treatment for low back pain in Chinese patients: a discrete choice experiment	PLoS One	2015
58	Chilton F, Collett RA	Treatment choices, preferences and decision-making by patients with rheumatoid arthritis	Musculoskeletal Care	2008
59	Chilvers C, Dewey M, Fielding K, et al	Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms	Bmj	2001

60*	Chuma J, Okungu V, Molyneux C	Barriers to prompt and effective malaria treatment among the poorest population in Kenya	Malaria Journal	2010
61	Chung IW, Kim YS, Lee NY, et al	Preference for long-acting injectable antipsychotics of community-dwelling patients with schizophrenia and their caregivers in Korea	International Journal of Neuropsychopharmacology	2014
62*	Chung VCH, Ma PHX, Lau CH, et al	Developing policy for integrating biomedicine and traditional Chinese medical practice using focus groups and the Delphi technique	Evidence-based Complementary and Alternative Medicine	2012
63	Cimminiello C, Anderson FA, Jr	Physician and patient perceptions of the route of administration of venous thromboembolism prophylaxis: results from an international survey	Thromb Res	2012
64	Clark DO, Kroenke K, Callahan CM, et al	Validity and utility of patient-reported health measures on hospital admission	J Clin Epidemiol	1999
65	Clatworthy J, Bowskill R, Rank T, et al	Adherence to medication in bipolar disorder: a qualitative study exploring the role of patients' beliefs about the condition and its treatment	Bipolar Disord	2007
66	Constantinescu F, Goucher S, Weinstein A, et al	Understanding why rheumatoid arthritis patient treatment preferences differ by race	Arthritis Rheum	2009
67	D'Abramo F, Schildmann J, Vollmann J	Research participants' perceptions and views on consent for biobank research: a review of empirical data and ethical analysis	BMC Med Ethics	2015
68*	Daftary A, Padayatchi N, O'Donnell M	Preferential adherence to antiretroviral therapy over tuberculosis treatment: a qualitative study of drug-resistant TB/HIV co-infected patients in South Africa	Glob Public Health	2014
69*	Dana Lamond NW, Skedgel C, Rayson D, et al	Adjuvant denosumab for breast cancer: What efficacy in the D-CARE trial will translate into cost effectiveness?	Journal of Clinical Oncology	2014
70	Dhanani MT, Wells AE, Montgomery TL, et al	Patient attitudes regarding the use of power morcellation during laparoscopic hysterectomy	Gynecologic Oncology	2015
71	Dickinson R, Hamrosi K, Knapp P, et al	Written medicines information in the UK and Australia: Exploring patient opinions and preferences on tailoring medicine leaflets	International Journal of Pharmacy Practice	2011

72	Dickinson R, Hamrosi K, Knapp P, et al	Suits you? A qualitative study exploring preferences regarding the tailoring of consumer medicines information	Int J Pharm Pract	2013
73	Dong D, Ozdemir S, Mong Bee Y, et al	Measuring High-Risk Patients' Preferences for Pharmacogenetic Testing to Reduce Severe Adverse Drug Reaction: A Discrete Choice Experiment	Value in Health	2016
74*	Doos L, Bradley E, Rushton CA, et al	Heart failure and chronic obstructive pulmonary disease multimorbidity at hospital discharge transition: a study of patient and carer experience	Health Expect	2015
75	Dowdy D, Bishai D, Chen AH	Setting clinical priorities: A framework for incorporating individual patient preferences	Patient Education and Counseling	2013
76*	Dowse R	Reflecting on patient-centred care in pharmacy through an illness narrative	Int J Clin Pharm	2015
77*	Dragojlovic N, Rizzardo S, Bansback N, et al	Challenges in measuring the societal value of orphan drugs: insights from a canadian stated preference survey	Patient	2015
78	Dranitsaris G, Leung P	Using decision modeling to determine pricing of new pharmaceuticals: The case of neurokinin-1 receptor antagonist antiemetics for cancer chemotherapy	International Journal of Technology Assessment in Health Care	2004
79	Duarte JW, Bolge SC, Sen SS	An evaluation of patients' preferences for osteoporosis medications and their attributes: the PREFER-International study	Clin Ther	2007
80	Dwight-Johnson M, Lagomasino IT, Aisenberg E, et al	Using conjoint analysis to assess depression treatment preferences among low-income Latinos	Psychiatr Serv	2004
81*	Emoto N, Okajima F, Sugihara H, et al	Behavioral economics survey of patients with type 1 and type 2 diabetes	Patient Preference and Adherence	2015
82	Ende J, Kazis L, Ash A, et al	Measuring patients' desire for autonomy - Decision-making and information-seeking preferences among medical patients	Journal of General Internal Medicine	1989
83*	Fargher E, Hughes D, Ring A, et al	Using discrete choice experiments to define patient preferences for outcomes in trials	Trials	2013

84	Feldman SR, Housman TS	Patients' vehicle preference for corticosteroid treatments of scalp psoriasis	Am J Clin Dermatol	2003
85*	Finckh A, Escher M, Liang MH, et al	Factors involved in the decision to take medications to prevent rheumatoid arthritis in first degree relatives of patients with rheumatoid arthritis: A discrete choice experiment	Arthritis and Rheumatism	2011
86	Fischer GS, Tulsky JA, Rose MR, et al	Patient knowledge and physician predictions of treatment preferences after discussion of advance directives	J Gen Intern Med	1998
87	Fraenkel L, Bogardus ST, Jr. , Concato J, et al	Treatment options in knee osteoarthritis: the patient's perspective	Arch Intern Med	2004
88	Fraenkel L, Gulanski B, Wittink D	Patient treatment preferences for osteoporosis	Arthritis Rheum	2006
89	Fraenkel L, Gulanski B, Wittink D	Patient willingness to take teriparatide	Patient Educ Couns	2007
90	Fraenkel L, Gulanski B, Wittink DR	Preference for hip protectors among older adults at high risk for osteoporotic fractures	J Rheumatol	2006
91*	Fukuda M, Suetsugu T, Ebi N, et al	Which do patientswith nsclc harboring EGFR mutation prefer EGFR-TKI or chemotherapy? A vignettes study (logik0903)	Annals of Oncology	2013
92	Fullerton L, Cesta A, Hofstetter C, et al	Patient priorities in ra research-an exploration of patient perspectives from those enrolled in the Ontario best practices research initiative (OBRI) clinical registry	Journal of Rheumatology	2015
93	Fyffe HE, Kay EJ	Assessment of dental health state utilities	Community Dent Oral Epidemiol	1992
94	Gale NK, Greenfield S, Gill P, et al	Patient and general practitioner attitudes to taking medication to prevent cardiovascular disease after receiving detailed information on risks and benefits of treatment: a qualitative study	BMC Fam Pract	2011
95*	Ganther JM, Wiederholt JB, Kreling DH	Measuring patients' medical care preferences: care seeking versus self-treating	Med Decis Making	2001
96*	Garrison LP, Jr., Towse A, Bresnahan BW	Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis	Health Aff (Millwood)	2007

97	Gauthier B, Singh SR, Virani A, et al	Perspectives and experiences of health care professionals and patients regarding treatments for type 2 diabetes	Canadian Pharmacists Journal	2014
98*	Gauthier C, Lindwall E, Davis W, et al	Spanning generations-appointment reminder preferences among patients with rheumatic diseases	J Clin Rheumatol	2012
99*	Gazzard B, Leen C, Morton J, et al	Evaluation of patient satisfaction with antiretroviral therapy using a discrete choice experiment	HIV Medicine	2015
100	Gelhorn HL, Bacci ED, Poon JL, et al	Evaluating preferences for profiles of glucagon-like peptide-1 receptor agonists among injection-naive type 2 diabetes patients in Japan	Patient Preference and Adherence	2016
101	Gelhorn HL, Poon JL, Davies EW, et al	Evaluating preferences for profiles of GLP-1 receptor agonists among injection-naïve type 2 diabetes patients in the UK	Patient Preference and Adherence	2015
102*	Gerard K, Salisbury C, Street D, et al	Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences	Journal of Health Services Research and Policy	2008
103	Gestuvo MK	Health maintenance in older adults: Combining evidence and individual preferences	Mount Sinai Journal of Medicine	2012
104*	Giebel GD, Groeben N	Social desirability in the measuring of patient satisfaction after treatment of coloproctologic disorders: on shortcomings of general bipolar satisfaction scales for quality management	Langenbecks Arch Surg	2008
105	Goertzel BN, Pennachin C, de Souza Coelho L, et al	Allostatic load is associated with symptoms in chronic fatigue syndrome patients	Pharmacogenomics	2006
106	Gold DT, Safi W, Trinh H	Patient preference and adherence: Comparative US studies between two bisphosphonates, weekly risedronate and monthly ibandronate	Current Medical Research and Opinion	2006
107	Gopal C, Ranga A, Joseph K, et al	Development and validation of algorithms for heart failure patient care: A Delphi study	International Journal of Pharmacy Practice	2014
108	Gould O, Buckley P, Doucette D	What patients want: Preferences regarding hospital pharmacy services	Canadian Journal of Hospital Pharmacy	2013

109	Grindrod KA, Marra CA, Colley L, et al	Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy	Ann Pharmacother	2010
110*	Grubbs V, Bibbins-Domingo K, Fernandez A, et al	Acute myocardial infarction length of stay and hospital mortality are not associated with language preference	Journal of General Internal Medicine	2008
111*	Guevara BEK, Gonzales N, Visitacion L	Patient preference on psoriasis treatment in a Philippine Tertiary Hospital: A conjoint analysis	Journal of the European Academy of Dermatology and Venereology	2016
112	Guirguis LM, Nusair MB	Standardized patients' preferences for pharmacist interactive communication style: A mixed method approach	Journal of the American Pharmacists Association	2016
113*	Gupta VK, Bahia JS, Maheshwari A, et al	To study the attitudes, beliefs and perceptions regarding the use of inhalers among patients of obstructive pulmonary diseases and in the general population in Punjab	Journal of Clinical and Diagnostic Research	2011
114	Hall J, Viney R, Haas M, et al	Using stated preference discrete choice modeling to evaluate health care programs	Journal of Business Research	2004
115	Hames H, Seabrook JA, Matsui D, et al	A palatability study of a flavored dexamethasone preparation versus prednisolone liquid in children	Journal canadien de pharmacologie clinique [Canadian journal of clinical pharmacology]	2008
116	Hamrosi K, Dickinson R, Knapp P, et al	It's for your benefit: exploring patients' opinions about the inclusion of textual and numerical benefit information in medicine leaflets	Int J Pharm Pract	2013
117	Hancock-Howard RL, Ungar WJ, Marshall D, et al	Public preferences for counseling regarding antidepressant use during pregnancy: a discrete choice experiment	Birth Defects Res A Clin Mol Teratol	2012
118*	Hanson KA, Agashivala N, Wyrwich KW, et al	Treatment selection and experience in multiple sclerosis: Survey of neurologists	Patient Preference and Adherence	2014
119	Harrison M, Marra C, Shojania K, et al	Societal preferences for rheumatoid arthritis treatments: evidence from a discrete choice experiment	Rheumatology (Oxford)	2015

120	Hart RI, McDonagh JE, Thompson B, et al	Being as Normal as Possible: How Young People Ages 16–25 Years Evaluate the Risks and Benefits of Treatment for Inflammatory Arthritis	Arthritis Care and Research	2016
121	Hayes RP, Bowman L, Monahan PO, et al	Understanding diabetes medications from the perspective of patients with type 2 diabetes: prerequisite to medication concordance	Diabetes Educ	2006
122	Hazlewood G, Marshall D, Tomlinson G, et al	Treatment preferences of patients with early rheumatoid arthritis: A discrete-choice experiment	Journal of Rheumatology	2015
123*	Heinze G, Cortés JF	Treatment preference and attitude toward pharmacotherapy and psychotherapy in Latin America. ULAD Task Force	Salud Mental	2005
124	Hess LM, Litwiller A, Byron J, et al	Preference elicitation tool for abnormal uterine bleeding treatment: a randomized controlled trial	Patient	2015
125*	Hill-Smith I, Mathie E, Little P	Involving patients in decisions about preventive medication: A focus group study	Primary Care Cardiovascular Journal	2010
126*	Hincapie A, Penm J, Burns C	A conjoint analysis of disease-modifying drugs used in the treatment of autoimmune conditions	Journal of the American Pharmacists Association	2016
127	Hissink MPCE, Brederije I, Brinkman DMC, et al	Parental preferences for treatment: Preliminary report from a randomised comparison of treatment strategies in (early) juvenile idiopathic arthritis (BeSt for Kids trial)	Pediatric Rheumatology	2011
128*	Hlubocky FJ, Ratain MJ, Wen M, et al	Complementary and alternative medicine among advanced cancer patients enrolled on phase I trials: a study of prognosis, quality of life, and preferences for decision-making	J Clin Oncol	2007
129	Hobson RJ, Scott J, Sutton J	Pharmacists and nurses as independent prescribers: exploring the patient's perspective	Fam Pract	2010
130*	Hoffmann C, Rosenberger A, Troeger W, et al	Validation of questionnaires from several medical fields regarding the constitution of patients	Forschende Komplementarmedizin und Klassische Naturheilkunde	2002

131	Holbrook K	A triangulation study of the clinician and patient experiences of the use of the immunosuppressant drugs azathioprine and 6-mercaptopurine for the management of inflammatory bowel disease	J Clin Nurs	2007
132*	Holden WL	Benefit-risk analysis : a brief review and proposed quantitative approaches	Drug Saf	2003
133	Hollin IL, Peay HL, Bridges JF	Caregiver preferences for emerging duchenne muscular dystrophy treatments: a comparison of Best-worst scaling and conjoint analysis	Patient	2015
134	Holmes EAF, Morrison VL, Hughes DA	What influences persistence with medicines? A multinational discrete choice experiment of 2549 patients	British Journal of Clinical Pharmacology	2016
135*	Hong SH, Liu J, Wang J, et al	Conjoint analysis of patient preferences on Medicare medication therapy management	J Am Pharm Assoc	2011
136*	Hood B, Egan R, Gridley H, et al	Treatment options for depression: Women and primary service providers	Australian Journal of Primary Health - Interchange	1999
137*	Housman TS, Mellen BG, Rapp SR, et al	Patients with psoriasis prefer solution and foam vehicles: a quantitative assessment of vehicle preference	Cutis cutaneous medicine for the practitioner	2002
138*	Hudak PL, Frankel RM, Braddock C, 3rd, et al	Do patients' communication behaviors provide insight into their preferences for participation in decision-making?	Med Decis Making	2008
139*	Hummel MJ, Volz F, van Manen JG, et al	Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment	Patient	2012
140*	Hur C, Broughton DE, Ozanne E, et al	Patient preferences for the chemoprevention of esophageal adenocarcinoma in Barrett's esophagus	Am J Gastroenterol	2008
141	Hussain-Gambles M	South Asian patients' views and experiences of clinical trial participation	Family Practice	2004

142	Ilozue C, Parirenyatwa R, Parris V, et al	Experience of switching patients from Atripla to Truvada and generic efavirenz in a medium-sized HIV clinic: Patient attitudes, safety and cost savings	HIV Medicine	2015
143*	Imaeda A, Bender D, Fraenkel L	What is most important to patients when deciding about colorectal screening?	Journal of General Internal Medicine	2010
144	Iskedjian M, Desjardins O, Piwko C, et al	Willingness to pay for a treatment for pain in multiple sclerosis	Pharmacoeconomics	2009
145*	Iskedjian M, Jaszewski B, Desjardins O	Comparison of scores derived from the box score-11 scale and the pain attribute of the health utilities index-3	Patient	2011
146	Iwama E, Takayama K, Suetsugu T, et al	Which do NSCLC patients with EGFR mutation prefer as a first-line therapy, EGFR-TKI or chemotherapy? A vignettes study (logik0903)	Respirology	2013
147	Jadoon NA, Sulehri FU, Shair NA, et al	Exploring patients' perspective and expectations regarding involvement in treatment decision-making	Asia-Pacific Journal of Clinical Oncology	2014
148	Jampel HD, Schwartz GF, Robin AL, et al	Patient preferences for eye drop characteristics: a willingness-to-pay analysis	Arch Ophthalmol	2003
149*	Jenkins V, Fallowfield L, Edginton T, et al	Preferences of healthy men for two different endocrine treatment options offered for locally advanced prostate cancer	Curr Med Res Opin	2005
150*	Jiang Q, Liu ZC, Zhang SX, et al	Age and cost are associated with patient preference for stopping tyrosine kinase inhibitor therapy in chronic myeloid leukemia	Blood	2015
151*	Jobanputra P, Wilson J, Douglas K, et al	A survey of British rheumatologists' DMARD preferences for rheumatoid arthritis	Rheumatology (Oxford)	2004
152	Johnson C, Lizama N, Saunders C	Australian general practitioners' preferences for inter-professional communication and a collaborative approach when caring for people diagnosed with cancer	Asia-Pacific Journal of Clinical Oncology	2012
153	Johnson ES, Sullivan SD, Mozaffari E, et al	A utility assessment of oral and intravenous ganciclovir for the maintenance treatment of AIDS-related cytomegalovirus retinitis	Pharmacoeconomics	1996

154	Johnson FR, Backhouse M	Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments	Value Health	2006
155	Joy SM, Little E, Maruthur NM, et al	Patient preferences for the treatment of type 2 diabetes: a scoping review	Pharmacoeconomics	2013
156*	Junaid R, Abaas M, Fatima B, et al	Attitude and practice of patients and doctors towards complementary and alternative medicine	J Pak Med Assoc	2012
157*	Kairuz TE, Jordaan I	Children's preferences for medicines - A pilot study in the Eastern Province region of South Africa	Journal of Applied Therapeutic Research	2004
158*	Karampli E, Triga E, Kyriopoulos J, et al	Views of physicians and patients with chronic conditions on generic medicines in Greece after the introduction of measures to promote their consumption: Findings from a qualitative study	GaBI Journal	2016
159	Karikios DJ, Howard K, Blinman P, et al	Oncologists' preferences for recommending expensive anticancer drugs	Asia-Pacific Journal of Clinical Oncology	2016
160	Kauf TL, Mohamed AF, Hauber AB, et al	Patients' willingness to accept the risks and benefits of new treatments for chronic hepatitis C virus infection	Patient	2012
161	Kawaguchi T, Azuma K, Yamaguchi T, et al	Preferences for pharmacist counselling in patients with breast cancer: A discrete choice experiment	Biological and Pharmaceutical Bulletin	2014
162	Kekwaletswe CT, Morojele NK	Alcohol use, antiretroviral therapy adherence, and preferences regarding an alcohol-focused adherence intervention in patients with human immunodeficiency virus	Patient Preference and Adherence	2014
163	Kellett N, West F, Finlay AY	Conjoint analysis: a novel, rigorous tool for determining patient preferences for topical antibiotic treatment for acne. A randomised controlled trial	Br J Dermatol	2006
164	Kennedy ED, To T, Steinhart AH, et al	Do patients consider postoperative maintenance therapy for Crohn's disease worthwhile?	Inflamm Bowel Dis	2008

165*	Kennelly C, Bowling A	Suffering in deference: a focus group study of older cardiac patients' preferences for treatment and perceptions of risk	Qual Health Care	2001
166	Kesternich I, Heiss F, McFadden D, et al	Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D	J Health Econ	2013
167	Kim MO, Lee KS, Kim JH, et al	Willingness to pay for hospice care using the contingent valuation method	Yonsei Medical Journal	2011
168	Kizilarslanoglu MC, Kilic MK, Toruk E, et al	Oral nutritional supplement taste preferences of inpatients	Gazi Medical Journal	2015
169*	Kleinman L, Kawata A, Harding G, et al	Patient willingness to pay for specific attributes of maintenance medication for COPD	American Journal of Respiratory and Critical Care Medicine	2010
170	Knass B, Empey P, Coley K	Patient attitudes toward pharmacogenomic testing in an independent community pharmacy	Journal of the American Pharmacists Association	2015
171	Kocher MS, Bishop J, Marshall R, et al	Operative versus nonoperative management of acute achilles tendon rupture: Expected-value decision analysis	American Journal of Sports Medicine	2002
172	Kopec JA, Richardson CG, Llewellyn-Thomas H, et al	Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied	J Clin Epidemiol	2007
173	Korthuis PT, Gregg J, Rogers WE, et al	Patients' reasons for choosing office-based buprenorphine: Preference for patient-centered care	Journal of Addiction Medicine	2010
174	Kovac JR, Rajanahally S, Smith RP, et al	Patient Satisfaction with Testosterone Replacement Therapies: The Reasons Behind the Choices	Journal of Sexual Medicine	2013
175	Kremer H, Ironson G	Measuring the involvement of people with HIV in treatment decision-making using the control preferences scale	Med Decis Making	2008
176*	Krishnan S, Gelhorn H, Merikle E, et al	A discrete choice experiment to elicit physician preferences for medication attributes in the treatment of patients with severe hemophilia A and inhibitors to factor VIII	Journal of Thrombosis and Haemostasis	2011

177*	Krüger K, Alten R, Schiffner-Rohe J, et al	Patient preferences in the choice of disease modifying anti-rheumatic drugs	Annals of the Rheumatic Diseases	2015
178	Kyle RMA, Lazo-Langner A, Xenocostas A, et al	Patient preferences for stopping tyrosine kinase inhibitors in chronic myeloid leukemia	Blood	2012
179	Laba TL, Brien JA, Jan S	Understanding rational non-adherence to medications. A discrete choice experiment in a community sample in Australia	BMC Fam Pract	2012
180	Laba TL, Essue B, Kimman M, et al	Understanding Patient Preferences in Medication Nonadherence: A Review of Stated Preference Data	Patient	2015
181*	Laba TL, Howard K, Rose J, et al	Using DCE to assess adherence and treatment preferences for combination therapies for cardiovascular disease	Global Heart	2014
182*	Laibovitz R, Boyle J, Snyder E, et al	Dorzolamide versus pilocarpine as adjunctive therapies to timolol: a comparison of patient preference and impact on daily life	Clin Ther	1996
183	Láinez-Andrés JM, Caminero AB, Díaz-Insa S, et al	Triptans in clinical practice: Efectiveness, tolerance and patient's satisfaction with medication (Tripracli study)	Revista de Neurologia	2013
184	Lancsar EJ, Hall JP, King M, et al	Using discrete choice experiments to investigate subject preferences for preventive asthma medication	Respirology	2007
185	Lee J, Pommerville P, Brock G, et al	Physician-rated patient preference and patient- and partner-rated preference for tadalafil or sildenafil citrate: results from the Canadian 'Treatment of Erectile Dysfunction' observational study	BJU Int	2006
186*	Lee JA, Nguyen AL, Berg J, et al	Attitudes and preferences on the use of mobile health technology and health games for self-management: Interviews with older adults on anticoagulation therapy	Journal of Medical Internet Research	2014
187*	Leonardi A, Zafirakis P	Efficacy and comfort of olopatadine versus ketotifen ophthalmic solutions: a double-masked, environmental study of patient preference	Curr Med Res Opin	2004

188*	Levitan B, Phillips LD, Walker S	Structured Approaches to Benefit-Risk Assessment: A Case Study and the Patient Perspective	Therapeutic Innovation and Regulatory Science	2014
189	Liang D, Tang CX	The specialty choice of medical students in China: A stated preference experiment	BMC Medical Education	2016
190*	Lichtenstein GR, Waters HC, Kelly J, et al	Assessing drug treatment preferences of patients with Crohn's disease: A conjoint analysis	Patient	2010
191	Liebrenz M, Gehring MT, Buadze A, et al	High-dose benzodiazepine dependence: a qualitative study of patients' perception on cessation and withdrawal	BMC Psychiatry	2015
192	Linley WG, Hughes DA	Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity	Pharmacoeconomics	2013
193	Liu H, Massi L, Laba TL, et al	Patients' and providers' perspectives of a polypill strategy to improve cardiovascular prevention in Australian Primary Health Care	Circulation: Cardiovascular Quality and Outcomes	2015
194	Lloyd A, McIntosh E, Price M	The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment	Pharmacoeconomics	2005
195	Lo PS, Tong MC, Revicki DA, et al	Rhinitis Symptom Utility Index (RSUI) in Chinese subjects: a multiattribute patient-preference approach	Qual Life Res	2006
196*	Loder E, Brandes JL, Silberstein S, et al	Preference comparison of rizatriptan ODT 10-mg and sumatriptan 50-mg tablet in migraine	Headache	2001
197*	Logge C, Vettorazzi E, Fischer L, et al	Antiviral treatment of HCV recurrence after liver transplantation: Impact on costs and quality of life	Transplantation	2010
198	Magin PJ, Adams J, Heading GS, et al	Complementary and alternative medicine therapies in acne, psoriasis, and atopic eczema: results of a qualitative study of patients' experiences and perceptions	J Altern Complement Med	2006
199	Mahadevia P, Shah S, Mannix S, et al	Willingness to pay for sensory attributes of intranasal corticosteroids among patients with allergic rhinitis	J Manag Care Pharm	2006

200	Mansfield C, Srinivas S, Chen C, et al	The effect of information on preferences for treatments of metastatic renal cell carcinoma	Oncology	2016
201	Man-Son-Hing M, Laupacis A, O'Connor A, et al	Warfarin for atrial fibrillation. The patient's perspective	Arch Intern Med	1996
202*	Marquis P, Arnould B, Acquadro C, et al	Patient-reported outcomes and health-related quality of life in effectiveness studies: Pros and cons	Drug Development Research	2006
203*	Matlock DD, Masoudi FA, Emsermann C, et al	What is more important in cardiologists' decision-making around implantable-cardioverter defibrillators (ICD), mortality data or patient preferences?	Circulation: Cardiovascular Quality and Outcomes	2011
204	Mauriello JA, Jr. , Dhillon S, Leone T, et al	Treatment selections of 239 patients with blepharospasm and Meige syndrome over 11 years	Br J Ophthalmol	1996
205	Mazur DJ	How older patient preferences are influenced by consideration of future health outcomes	J Am Geriatr Soc	1997
206*	Mazur DJ, Merz JF	How the manner of presentation of data influences older patients in determining their treatment preferences	J Am Geriatr Soc	1993
207*	Mazur DJ, Merz JF	How older patients' treatment preferences are influenced by disclosures about therapeutic uncertainty: surgery versus expectant management for localized prostate cancer	J Am Geriatr Soc	1996
208	McCarthy AE, Coyle D	Determining utility values related to malaria and malaria chemoprophylaxis	Malar J	2010
209	McCombs K, Chen SC	Patient preference quality of life measures in dermatology	Dermatol Ther	2007
210*	McLeod C, Bagust A, Boland A, et al	Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation	Health Technol Assess	2007
211	McSharry J, McGowan L, Farmer AJ, et al	Perceptions and experiences of taking oral medications for the treatment of Type 2 diabetes mellitus: a systematic review and meta-synthesis of qualitative studies	Diabetic Medicine	2016

212	McTaggart-Cowan HM, Shi P, Fitzgerald JM, et al	An evaluation of patients' willingness to trade symptom-free days for asthma-related treatment risks: a discrete choice experiment	J Asthma	2008
213	Mele NL	Conjoint analysis: using a market-based research model for healthcare decision-making	Nurs Res	2008
214*	Metlay JP, Shea JA, Crossette LB, et al	Tensions in antibiotic prescribing: pitting social concerns against the interests of individual patients	J Gen Intern Med	2002
215	Mohamed AF, Johnson FR, Hauber AB, et al	Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey, and Italy	Eur J Gastroenterol Hepatol	2012
216	Mohsin-Shaikh S, Garfield S, Franklin BD	Patient involvement in medication safety in hospital: an exploratory study	Int J Clin Pharm	2014
217	Mooney K, McElnay JC, Donnelly RF	Parents' perceptions of microneedle-mediated monitoring as an alternative to blood sampling in the monitoring of their infants	International Journal of Pharmacy Practice	2015
218	Morel T, Aymé S, Cassiman D, et al	Quantifying benefit-risk preferences for new medicines in rare disease patients and caregivers	Orphanet Journal of Rare Diseases	2016
219	Morss SE, Lenert LA, Faustman WO	The side effects of antipsychotic drugs and patients' quality of life: patient education and preference assessment with computers and multimedia	Proc Annu Symp Comput Appl Med Care	1993
220	Muduma G, Shupo FC, Dam S, et al	Patient survey to identify reasons for non-adherence and elicitation of quality of life concepts associated with immunosuppressant therapy in kidney transplant recipients	Patient Preference and Adherence	2016
221	Mühlbacher AC, Bethge S, Kaczynski A	TREATMENT after ACUTE CORONARY SYNDROME: ANALYSIS of PATIENT'S PRIORITIES with ANALYTIC HIERARCHY PROCESS	International Journal of Technology Assessment in Health Care	2016
222	Mühlbacher AC, Ñbling M	Analysis of patients preferences: Direct assessment and discrete-choice experiment in therapy of adults with attention-deficit hyperactivity disorder	Patient	2010

223	Müller-Ladner U, Rockwitz K, Brandt-Jürgens J, et al	Tolerability and patient/physician satisfaction with subcutaneously administered methotrexate provided in two formulations of different drug concentrations in patients with rheumatoid arthritis	Open Rheumatology Journal	2010
224	Müller-Schwefe GHH, Wimmer A, Eggers A, et al	Patient's perspectives on opioid dosing schedules for chronic pain in Germany, Italy, and Turkey: Pooled analysis from pares study	European Journal of Pain Supplements	2011
225*	Murphy DJ, Gahm GJ, Santilli S, et al	Seniors' preferences for cancer screening and medication use based on absolute risk reduction	J Gerontol A Biol Sci Med Sci	2002
226	Naik Panvelkar P, Armour C, Saini B	Community pharmacy-based asthma services-what do patients prefer	Journal of Asthma	2010
227	Najafzadeh M, Johnston KM, Peacock SJ, et al	Genomic testing to determine drug response: measuring preferences of the public and patients using Discrete Choice Experiment (DCE)	BMC Health Serv Res	2013
228*	Nexoe J, Kristiansen IS, Gyrd-Hansen D, et al	Influence of number needed to treat, costs and outcome on preferences for a preventive drug	Fam Pract	2005
229	Ngorsuraches S, Thongkeaw K	Breast cancer patients' preferences and willingness to pay for mtor inhibitors	Journal of the American Pharmacists Association	2015
230	Niemand AM, Bausewein C, Voltz R, et al	Acceptability and preferences of six different modes of drug application for acute episodes of breathlessness (EB)-a comparison study between two countries	Palliative Medicine	2012
231	Nolting SK, Sanchez Carazo S, De Boulle K, et al	Oral treatment schedules for onychomycosis: a study of patient preference	Int J Dermatol	1998
232	Nordon C, Falissard B, Gerard S, et al	Patient satisfaction with psychotropic drugs: Validation of the PAtient SAtisfaction with Psychotropic (PASAP) scale in patients with bipolar disorder	Eur Psychiatry	2014

233*	O'Connor AM, Boyd NF, Tritchler DL, et al	Eliciting preferences for alternative cancer drug treatments. The influence of framing, medium, and rater variables	Medical decision-making : an international journal of the Society for Medical Decision-making	1985
234	O'Connor AM, Boyd NF, Warde P, et al	Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome description, elicitation technique and treatment experience on preferences	Journal of Chronic Diseases	1987
235	O'Donoghue B, Lyne J, Hill M, et al	Patient attitudes towards compulsory community treatment orders and advance directives	Irish Journal of Psychological Medicine	2010
236	Pagès-Puigdemont N, Mangues MA, Masip M, et al	Patients' Perspective of Medication Adherence in Chronic Conditions: A Qualitative Study	Advances in Therapy	2016
237	Palmer CS, Schmier JK, Snyder E, et al	Patient preferences and utilities for "off-time" outcomes in the treatment of Parkinson's disease	Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation	2000
238	Park H, Chon Y, Lee J, et al	Service design attributes affecting diabetic patient preferences of telemedicine in South Korea	Telemedicine and e Health	2011
239	Park Y, Lee JY, Nam EJ, et al	Treatment preferences of patients with advanced ovarian cancer for the addition of bevacizumab to first line therapy	International Journal of Gynecological Cancer	2016
240*	Partridge B, Lucke J, Hall W	Over-diagnosed and over-treated: A survey of Australian public attitudes towards the acceptability of drug treatment for depression and ADHD	BMC Psychiatry	2014
241*	Patel SR, Simpson HB	Patient preferences for obsessive-compulsive disorder treatment	J Clin Psychiatry	2010
242*	Peyrot M, Rubin RR	How does treatment satisfaction work?: Modeling determinants of treatment satisfaction and preference	Diabetes Care	2009
243	Pignone MP, Brenner AT, Hawley S, et al	Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening	Journal of general internal medicine	2012

244	Poder TG, He J, Simard C, et al	Willingness to pay for ovulation induction treatment in case of WHO II anovulation: A study using the contingent valuation method	Patient Preference and Adherence	2014
245*	Postmus D, Mavris M, Hillege HL, et al	Incorporating patient preferences into drug development and regulatory decision-making: Results from a quantitative pilot study with cancer patients, carers, and regulators	Clin Pharmacol Ther	2016
246	Poulos C, Kinter E, Yang JC, et al	A discrete-choice experiment to determine patient preferences for injectable multiple sclerosis treatments in Germany	Therapeutic Advances in Neurological Disorders	2016
247	Powell G, Holmes EA, Plumpton CO, et al	Pharmacogenetic testing prior to carbamazepine treatment of epilepsy: patients' and physicians' preferences for testing and service delivery	Br J Clin Pharmacol	2015
248*	Prothero L, Georgopoulou S, Galloway J, et al	Patients' and carers' views and expectations about intensive management for moderate rheumatoid arthritis: A qualitative study	Psychology, Health & Medicine	2016
249	Quaife M, Eakle R, Cabrera M, et al	Preferences for ARV-based HIV prevention methods among men and women, adolescent girls and female sex workers in Gauteng Province, South Africa: A protocol for a discrete choice experiment	BMJ Open	2016
250*	Quinn J, Doll H, Lewis H, et al	Determining treatment preference for tolvaptan in autosomal dominant polycystic kidney disease (ADPKD): Development of a discrete choice experiment (DCE) for use as a clinical study endpoint	Nephrology Dialysis Transplantation	2016
251	Ranmal S, Tuleu C	Piloting the 'Children's acceptability of oral formulations (CALF) medicines survey to determine perceptions and practices with paediatrics medicines	Archives of Disease in Childhood	2013
252*	Rapoport AM, Tepper SJ, Bigal ME, et al	The triptan formulations : how to match patients and products	CNS Drugs	2003
253	Robberstad B, Olsen JA	The health related quality of life of people living with HIV/AIDS in sub-Saharan Africa - a literature review and focus group study	Cost Effectiveness and Resource Allocation	2010

254*	Ross JD, Copas A, Stephenson J, et al	Public involvement in modernising genitourinary medicine clinics: using general public and patient opinion to influence models of service delivery	Sex Transm Infect	2006
255	Rowe G, Lambert N, Bowling A, et al	Assessing patients' preferences for treatments for angina using a modified repertory grid method	Soc Sci Med	2005
256*	Salas M, Rosas M, Pastelin G	Patient preferences for antihypertensive medications in general practice	Pharmacoepidemiology and Drug Safety	2013
257	Salloum RG, Maziak W, Hammond D, et al	Eliciting preferences for waterpipe tobacco smoking using a discrete choice experiment: implications for product regulation	BMJ Open	2015
258*	Sattar T, Jadoonanan H, Ledbetter S	The utility of the medical home: a survey on patient perspectives	Osteopathic Family Physician	2010
259*	Schalkwyk J, Amiri N, Lalji S, et al	Acceptance of a rapid herpes test in labour: survey of attitudes of patients and health care providers	J Obstet Gynaecol Can	2008
260	Schluter PJ, Ware RS	Single patient (n-of-1) trials with binary treatment preference	Statistics in Medicine	2005
261*	Schumm JA, Walter KH, Bartone AS, et al	Veteran satisfaction and treatment preferences in response to a posttraumatic stress disorder specialty clinic orientation group	Behaviour Research and Therapy	2015
262	Schwartz CE, Merriman MP, Reed GW, et al	Measuring patient treatment preferences in end-of-life care research: applications for advance care planning interventions and response shift research	J Palliat Med	2004
263*	Scott-Horton T, Vest K, Kliethermes MA	Using patient preference to customize a Patient's medication list	Consultant Pharmacist	2014
264	Shanahan M, Gerard K, Ritter A	Preferences for policy options for cannabis in an Australian general population: A discrete choice experiment	Int J Drug Policy	2014
265*	Sheftell FD, Fox AW	Acute migraine treatment outcome measures: a clinician's view	Cephalalgia	2000
266	Shingler SL, Swinburn P, Ali S, et al	A discrete choice experiment to determine patient preferences for injection devices in multiple sclerosis	J Med Econ	2013

267	Shiyanbola O, Ghura S, Huang Y, et al	Refining empirically designed prescription warning labels using patient feedback: A qualitative study	Journal of the American Pharmacists Association	2016
268	Shiyanbola O, Ghura S, Huang Y, et al	Pharmacist and patient feedback on empirically designed prescription warning labels: Similarities and differences	Journal of the American Pharmacists Association	2016
269	Shuford K, Were F, Awino N, et al	Community perceptions of mass screening and treatment for malaria in Siaya County, western Kenya	Malaria Journal	2016
270*	Siegel CA	Shared decision-making in inflammatory bowel disease: helping patients understand the tradeoffs between treatment options	Gut	2012
271*	Siegel CA, Levy LC, Mackenzie TA, et al	Patient perceptions of the risks and benefits of infliximab for the treatment of inflammatory bowel disease	Inflamm Bowel Dis	2008
272	Silverman S, Calderon A, Kaw K, et al	Patient weighting of osteoporosis medication attributes across racial and ethnic groups: a study of osteoporosis medication preferences using conjoint analysis	Osteoporos Int	2013
273*	Simon JA, Lewiecki EM, Smith ME, et al	Patient preference for once-weekly alendronate 70 mg versus once-daily alendronate 10 mg: a multicenter, randomized, open-label, crossover study	Clinical therapeutics	2002
274	Singh J, Longworth L, Baine A, et al	Exploring what lies behind public preferences for avoiding health losses caused by lapses in healthcare safety and patient lifestyle choices	BMC Health Serv Res	2013
275*	Sobo EJ, Billman G, Lim L, et al	A rapid interview protocol supporting patient-centered quality improvement: hearing the parent's voice in a pediatric cancer unit	The Joint Commission journal on quality improvement	2002
276	Stan DL, Pruthi S, Jenkins S, et al	Needs and preferences of breast cancer survivors: A cross-sectional survey	Journal of Clinical Oncology	2011
277	Stavem K	Quality of life in epilepsy: comparison of four preference measures	Epilepsy Res	1998

278	Stockler M, Vardy J, Pillai A, et al	Acetaminophen (paracetamol) improves pain and well-being in people with advanced cancer already receiving a strong opioid regimen: a randomized, double-blind, placebo-controlled cross-over trial	J Clin Oncol	2004
279	Stroberg P, Murphy A, Costigan T	Switching patients with erectile dysfunction from sildenafil citrate to tadalafil: results of a European multicenter, open-label study of patient preference	Clin Ther	2003
280	Sugar CA, James GM, Lenert LA, et al	Discrete state analysis for interpretation of data from clinical trials	Med Care	2004
281	Szeinbach SL, Barnes JH, McGhan WF, et al	Using conjoint analysis to evaluate health state preferences	Drug Information Journal	1999
282	Tallon D, Chard J, Dieppe P	Relation between agendas of the research community and the research consumer	Lancet	2000
283	Tallon D, Chard J, Dieppe P	Exploring the priorities of patients with osteoarthritis of the knee	Arthritis Care Res	2000
284	Tang ST, Wen FH, Hsieh CH, et al	Preferences for life-sustaining treatments and associations with accurate prognostic awareness and depressive symptoms in terminally ill cancer patients' last year of life	Journal of Pain and Symptom Management	2016
285	Tarn DM, Guzman JR, Good JS, et al	Provider and patient expectations for dietary supplement discussions	J Gen Intern Med	2014
286	Teasdale E, Santer M, Geraghty AW, et al	Public perceptions of non-pharmaceutical interventions for reducing transmission of respiratory infection: systematic review and synthesis of qualitative studies	BMC Public Health	2014
287	Thirumurthy H, Hayashi K, Linnemayr S, et al	Time Preferences Predict Mortality among HIV-Infected Adults Receiving Antiretroviral Therapy in Kenya	PLoS One	2015
288	Thomasius F, Keung NT, Ivan P	Phase IV randomized preference study in patients eligible for calcium and vitamin D supplementation	Current medical research and opinion	2016
289*	Thorpe CT, DeVellis RF, Blalock SJ, et al	Patient perceptions about illness self-management in ANCA-associated small vessel vasculitis	Rheumatology (Oxford)	2008

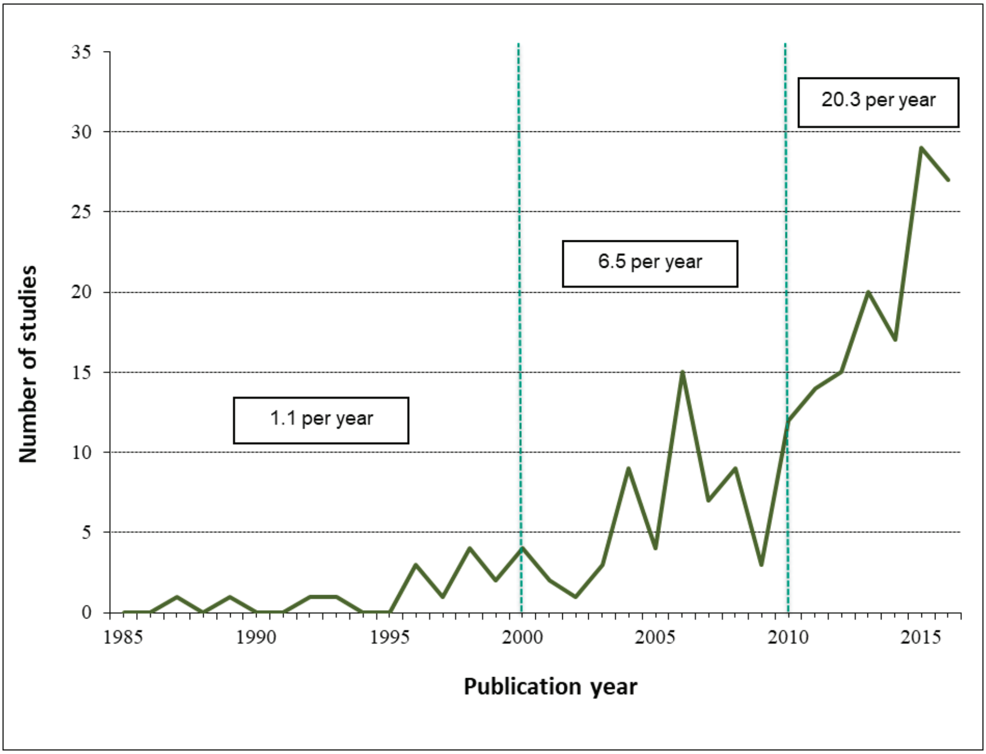
290	Tong V, Raynor DK, Blalock SJ, et al	Exploring consumer opinions on the presentation of side-effects information in Australian consumer medicine information leaflets	Health Expectations: An International Journal of Public Participation in Health Care & Health Policy	2016
291*	Tony M, Wagner M, Khoury H, et al	Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada	BMC Health Serv Res	2011
292*	Torres-Vigil I, De La Rosa A, Peña E, et al	End-of-life preferences of older latinos with advanced cancer and their community-dwelling contemporaries and consistency between patient preferences and care	Palliative Medicine	2014
293*	Toscano D, Brice J, Alfaro C	Usage and perceptions of pen injectors for diabetes management: a survey of type 2 diabetes patients in the United States	J Diabetes Sci Technol	2012
294	Treloar C, Newland J, Maher L	A qualitative study trialling the acceptability of new hepatitis C prevention messages for people who inject drugs: symbiotic messages, pleasure and conditional interpretations	Harm Reduct J	2015
295*	Trujols J, Iraurgi I, Sinol N, et al	Satisfaction with methadone as a medication: psychometric properties of the Spanish version of the treatment satisfaction questionnaire for medication	J Clin Psychopharmacol	2012
296*	Tsou A, Long J, McCluskey L, et al	Measuring value: Patient preferences for riluzole in amyotrophic lateral sclerosis (ALS)	Neurology	2012
297*	Tudiver F, Wolff LT, Morin PC, et al	Primary care providers' perceptions of home diabetes telemedicine care in the IDEATel project	J Rural Health	2007
298	Ubel PA, Spranca MD, Dekay ML, et al	Public preferences for prevention versus cure: What if an ounce of prevention is worth only an ounce of cure?	edical Decision-making	1998

299*	Underwood M, Ashby D, Carnes D, et al	Topical or oral ibuprofen for chronic knee pain in older people. The TOIB study	Health Technol Assess	2008
300	Ungar WJ, Hadioonzadeh A, Najafzadeh M, et al	Parents and adolescents preferences for asthma control: a Best-worst scaling choice experiment using an orthogonal main effects design	BMC Pulm Med	2015
301	Uribe A, Tucker LB, Masse L, et al	Parents' attitudes towards research and research study participation in pediatric rheumatology (PR)	Arthritis and Rheumatism	2011
302	Utens CMA, Dirksen CD, van der Weijden T, et al	How to integrate research evidence on patient preferences in pharmaceutical coverage decisions and clinical practice guidelines: A qualitative study among Dutch stakeholders	Health Policy	2016
303	van MH, Allick G, Wegman F, et al	Alprazolam for depression	Cochrane Database of Systematic Reviews	2012
304	van Summeren JJGT, Haaijer-Ruskamp FM, Schuling J	Eliciting Preferences of Multimorbid Elderly Adults in Family Practice Using an Outcome Prioritization Tool	Journal of the American Geriatrics Society	2016
305*	van Walraven C, Duke SM, Weinberg AL, et al	Standardized or narrative discharge summaries. Which do family physicians prefer?	Can Fam Physician	1998
306	Vina ER, Masi CM, Green SL, et al	A study of racial/ethnic differences in treatment preferences among lupus patients	Rheumatology (Oxford)	2012
307	Vina ER, Utset TO, Hannon MJ, et al	Racial differences in treatment preferences among lupus patients: a two-site study	Clin Exp Rheumatol	2014
308	Vincent C	Developing a policy for integrating biomedicine and traditional Chinese medical practice using focus groups and the Delphi technique	European Journal of Integrative Medicine	2012
309	Wang J, Hong SH, Meng S, et al	Pharmacists' acceptable levels of compensation for MTM services: a conjoint analysis	Res Social Adm Pharm	2011
310	White JH, Towers SE, Turner A, et al	Electronic screening and decision support for poststroke depression: An exploration of doctors' and patients' perceptions of acceptability	Archives of Physical Medicine and Rehabilitation	2013

311	Whitty JA, Rundle-Thiele SR, Scuffham PA	Insights from triangulation of two purchase choice elicitation methods to predict social decision-making in healthcare	Appl Health Econ Health Policy	2012
312	Wilson L, Loucks A, Bui C, et al	Patient centered decision-making: use of conjoint analysis to determine risk-benefit trade-offs for preference sensitive treatment choices	J Neurol Sci	2014
313	Wittink MN, Cary M, TenHave T, et al	Towards patient-centered care for depression: Conjoint methods to tailor treatment based on preferences	Patient	2010
314	Wolfe F, Zhao S, Lane N	Preference for nonsteroidal antiinflammatory drugs over acetaminophen by rheumatic disease patients: a survey of 1,799 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia	Arthritis Rheum	2000
315*	Wong A, Kraus PS, Lau BD, et al	Patient preferences regarding pharmacologic venous thromboembolism prophylaxis	J Hosp Med	2015
316	Woolf SH, Krist AH, Johnson RE, et al	Unwanted control: how patients in the primary care setting decide about screening for prostate cancer	Patient Educ Couns	2005
317	Wu F, Peng CY, Jiang H, et al	Methadone maintenance treatment in China: perceived challenges from the perspectives of service providers and patients	J Public Health (Oxf)	2013
318	Wu JM, Fulton RG, Amundsen CL, et al	Patient preferences for different severities of and treatments for overactive bladder	Female Pelvic Medicine and Reconstructive Surgery	2011
319*	Yarris LM, Moreno R, Schmidt TA, et al	Reasons why patients choose an ambulance and willingness to consider alternatives	Academic Emergency Medicine	2006
320	Zafar AM, Harris TJ, Murphy TP, et al	Patients' perspective about risks and benefits of treatment for peripheral arterial disease	J Vasc Interv Radiol	2011
321*	Zhang G, Parikh PB, Zabihi S, et al	Rating the preferences for potential harms of treatments for cardiovascular disease: a survey of community-dwelling adults	Med Decis Making	2013
322*	Zipursky RB, Cunningham CE, Bieling P, et al	Evaluating patient preferences for early intervention services using discrete choice conjoint analysis	Schizophrenia Research	2012
323	Zuchowski JL, Hamilton AB, Pyne JM, et al	Qualitative analysis of patient-centered decision attributes associated with initiating hepatitis C treatment	BMC Gastroenterol	2015

*Excluded studies for data extraction through full-text analysis

Appendix C: Number of patient preference studies within the medical product lifecycle per publication year, including means for different time periods



Chapter 3

Discrete Choice Experiments in Health Economics: Past, Present and Future

Soekhai V, de Bekker-Grob EW, Ellis AR and Vass CM

Published PharmacoEconomics 2019

Abstract

Objectives: Discrete choice experiments (DCEs) are increasingly advocated as a way to quantify preferences for health. However, increasing support does not necessarily result in increasing quality. Although specific reviews have been conducted in certain contexts, there exists no recent description of the general state of the science of health-related DCEs. The aim of this paper was to update prior reviews (1990-2012), to identify all health-related DCEs and to provide a description of trends, current practice and future challenges.

Methods: A systematic literature review was conducted to identify health-related empirical DCEs published between 2013-2017. The search strategy and data extraction replicated prior reviews to allow the reporting of trends, although additional extraction fields were incorporated.

Results: Of the 7877 abstracts generated, 301 studies met the inclusion criteria and underwent data extraction. In general, the total number of DCEs per year continued to increase, with broader areas of application and increased geographic scope. Studies reported using more sophisticated designs (e.g. D-efficient) with associated software (e.g. Ngene). The trend towards using more sophisticated econometric models also continued. However, many studies presented sophisticated methods with insufficient detail. Qualitative research methods continued to be a popular approach for identifying attributes and levels.

Conclusions: The use of empirical DCEs in health economics continues to grow. However, inadequate reporting of methodological details inhibits quality assessment. This may reduce decision-makers' confidence in results and their ability to act on the findings. How and when to integrate health-related DCE outcomes into decision-making remains an important area for future research.

3.1 Introduction

In recent years, there have been increased calls for patient and public involvement in healthcare decision-making [1-2]. Patient or public involvement can support decision-making at multiple levels: individual (shared decision-making), policy (patient experts on panels) and commissioning (incorporating patient preferences in technology evaluations or health state valuation). Views can be elicited qualitatively, quantitively or using mixed-methods approaches [3]. Example methods include interviews, focus groups and stated preference techniques such as the standard gamble or time trade-off. Studies by the Medical Device Innovation Consortium (MDIC) [4] and Mahieu *et al*. [5] highlighted a wide variety of methods to measure both stated and revealed preferences in healthcare.

Among the quantitative methods for eliciting stated health preferences, discrete choice experiments (DCEs) are increasingly advocated [6]. In a DCE individuals are asked to select their preferred (and/or least preferred) alternative from a set of alternatives. DCEs are grounded in theories which assume that: (1) alternatives can be described by their attributes, (2) an individual's valuation depends upon the levels of these attributes and (3) choices are based on a latent utility function [7-10]. The theoretical foundations have implications for the experimental design (principles to construct alternatives and choice sets) and the probabilistic models used to analyse the choice data [7].

Previously conducted broad reviews by Ryan and Gerard (1990-2000) [11], de Bekker-Grob *et al.* (2001-2008) [7] and Clark *et al.* (2009-2012) [6] identified a number of methodological challenges of DCEs (e.g. how to choose among orthogonal, D-efficient and other designs or how to account for preference heterogeneity when analysing choice data). These reviews, as well as published checklists [12] and best-practice guidelines [13-17], have been developed to provide specific guidance and potentially improve quality [18-19]. However, it is unknown whether the challenges identified in prior reviews are still relevant or whether there has been a response to the published suggestions and guidelines. Furthermore, although health-related DCEs are increasingly advocated by organisations such as the MDIC [4], their use for actual decision-making in health remains limited [7,13]. Key barriers to their wider use in policy include concerns about the robustness and validity of the method and the quality of applied studies [20,21].

This paper seeks to provide a current overview of the applications and methods used by DCEs in health economics. This overview will be created by systematically reviewing DCE literature and extracting data from the period from 2013–2017. In addition, historical trends in experimental design, analytical methods, validation procedures and outcome measures will be described by comparing the results to those of prior reviews. For the sake of generality and to allow examination of trends based on consistent data extraction methods, this comparison will focus on the broad reviews cited above, rather than on narrower reviews of DCEs covering specific study designs or disease areas [22-40]. Recent developments in DCE methods will be incorporated by including new data elements not reported in previous reviews. Potential challenges and recommendations for future research will also be identified.

3.2 Methods

The current systematic review continued the work conducted in the prior broad DCE reviews [6,7,11] by focusing on DCE[iv] applications published between 2013 and 2017. The methodology for this systematic review built on that of the prior reviews to allow comparison of results across review periods and identification of trends. The search was initiated in May 2015 and updated in February 2016 and January 2018. We used the same search engine

[iv] In this review best-worst-scaling (BWS) case 1 and 2 are distinguished from case 3. Since case 1 and 2 BWS do not involve attribute-based comparisons between two or more alternatives, they were excluded from this review [365], consistent with the previous review [6]. Case 3 BWS, however, involves an attribute-based comparison between two or more alternatives and is considered an extension of DCEs in the literature [366,367]. Therefore, case 3 BWS applications were included in this review.

(PubMed) that was used in the latest review by Clark *et al.* [6] and generally used the same search terms. We decided to exclude the search terms ‘conjoint’ and ‘dce’, since these yielded too many irrelevant results (particularly due to the rise of dynamic contrast enhanced imaging in gene expression profiling) and would have substantially increased the number of abstracts to be reviewed. The final search terms included: ‘discrete choice experiment’, ‘discrete choice experiments’, ‘discrete choice modeling’, ‘discrete choice modelling’, ‘discrete choice conjoint experiment’, ‘stated preference’, ‘part-worth utilities’, ‘functional measurement’, ‘paired comparisons’, ‘pairwise choices’, ‘conjoint analysis’, ‘conjoint measurement’, ‘conjoint studies’, ‘conjoint choice experiment’ and ‘conjoint choice experiments’. A study was included if it was applied to health, included a discrete choice exercise (rather than rating or ranking), focused on human beings and was published as a full-text article in English between January 2013 and December 2017. Consistent with prior reviews, DCEs without empirical data (e.g. methodological studies) and studies of samples already included in our review were excluded.

To ensure consistency of data extraction and assist with synthesis of results, the authors used an extraction tool, available in Appendix A of the Electronic Supplementary Material, initially developed using the criteria of Clark *et al.* [6]. We first considered areas of application (e.g. patient consumer experience, valuing health outcomes) and background information (country of origin, number and type of attributes, number of choice sets, survey administration method), followed by more detailed information about the experimental design (type, plan, use of blocking, design software, design source, method used to create choice sets, number of alternatives, presence of an opt-out or status-quo option, sample size and type), data analysis (model, analysis software, model details), validity checks (external and internal), use of qualitative methods (type and rationale) and presented outcome measures. The authors tested the extraction tool and discussed initial results. To fully capture current

DCE design methods, the following data elements were added to the original data extraction tool: number of alternatives, presence of an opt-out or status quo, sample size, use of blocking, use of a Bayesian design approach, software for econometric analyses and the type of qualitative research methods reported. With regard to analysis methods, this review also extracted additional information on the use of scale adjusted latent class, heteroskedastic conditional logit and generalised multinomial models. Studies were also categorised by journal type.

Each author extracted data from a group of articles, checking online appendices and supplementary materials where relevant. A subsample of studies (20%) was double-checked by VS for quality control. We categorized the extracted data and reported the results as percentages. Results for the econometric analysis models were categorized based on the three key characteristics of the multinomial logit model (Figure 1): (i) the assumption that error terms are independent and identically distributed (IID) according to the Extreme Value type I distribution, (ii) independence of irrelevant alternatives (IIA) (resulting from the first characteristic) and (iii) the presence or absence of preference heterogeneity [7]. The IID characteristic limits flexibility in estimating the error variance, whereas IIA is about the flexibility of the substitution pattern (how flexible respondents are to substitute between choices) and assumptions about preference heterogeneity determine whether preferences are allowed to vary across respondents.

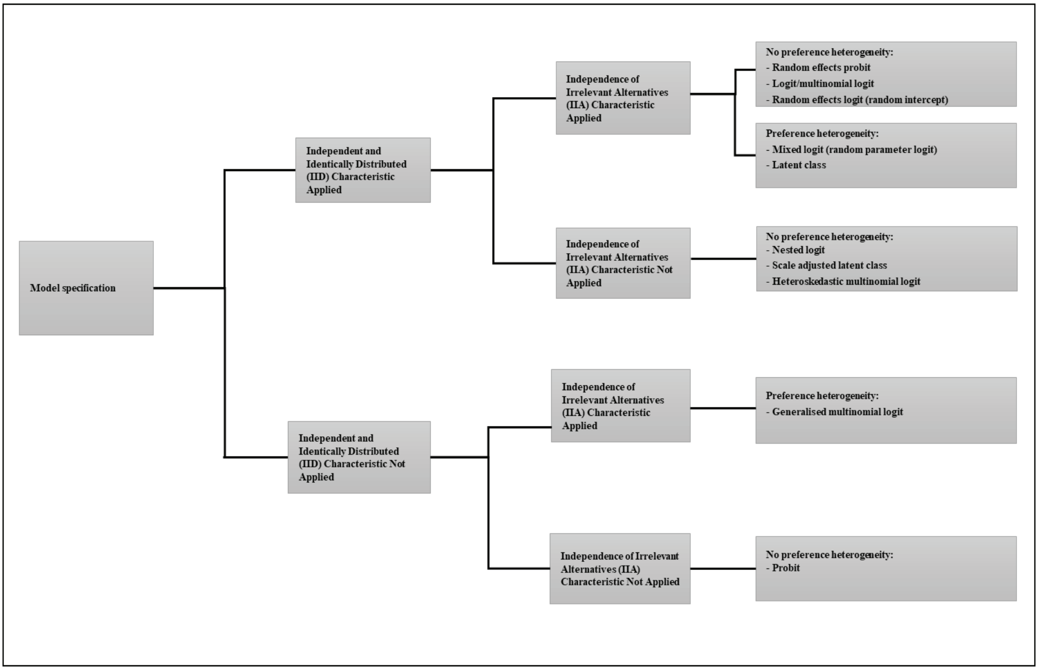


Fig. 1 Econometric analysis model overview

3.3 Results

3.3.1 Search results

A total of 7877 abstracts were identified from the beginning of 2013 until the end of 2017. After abstract and full-text review, 301 DCEs (including 6 case 3 BWS studies) met the inclusion criteria and were selected for data extraction (see Figure 2) [64-364]. Figure 3 depicts the total number of DCE applications in health across the different review periods 1990-2000, 2001-2008, 2009-2012 and 2013-2017. The 2009-2012 review reported that the number of studies had increased to 45 per year on average [6]. The current review period found 60 studies per year on average, with a high of 98 studies in 2015 and a low of 32 studies in 2017 (Figure 3). Figure 3 also shows that the increase in DCE applications between the prior review periods and the current review period was less consistent than the increases observed in prior periods.

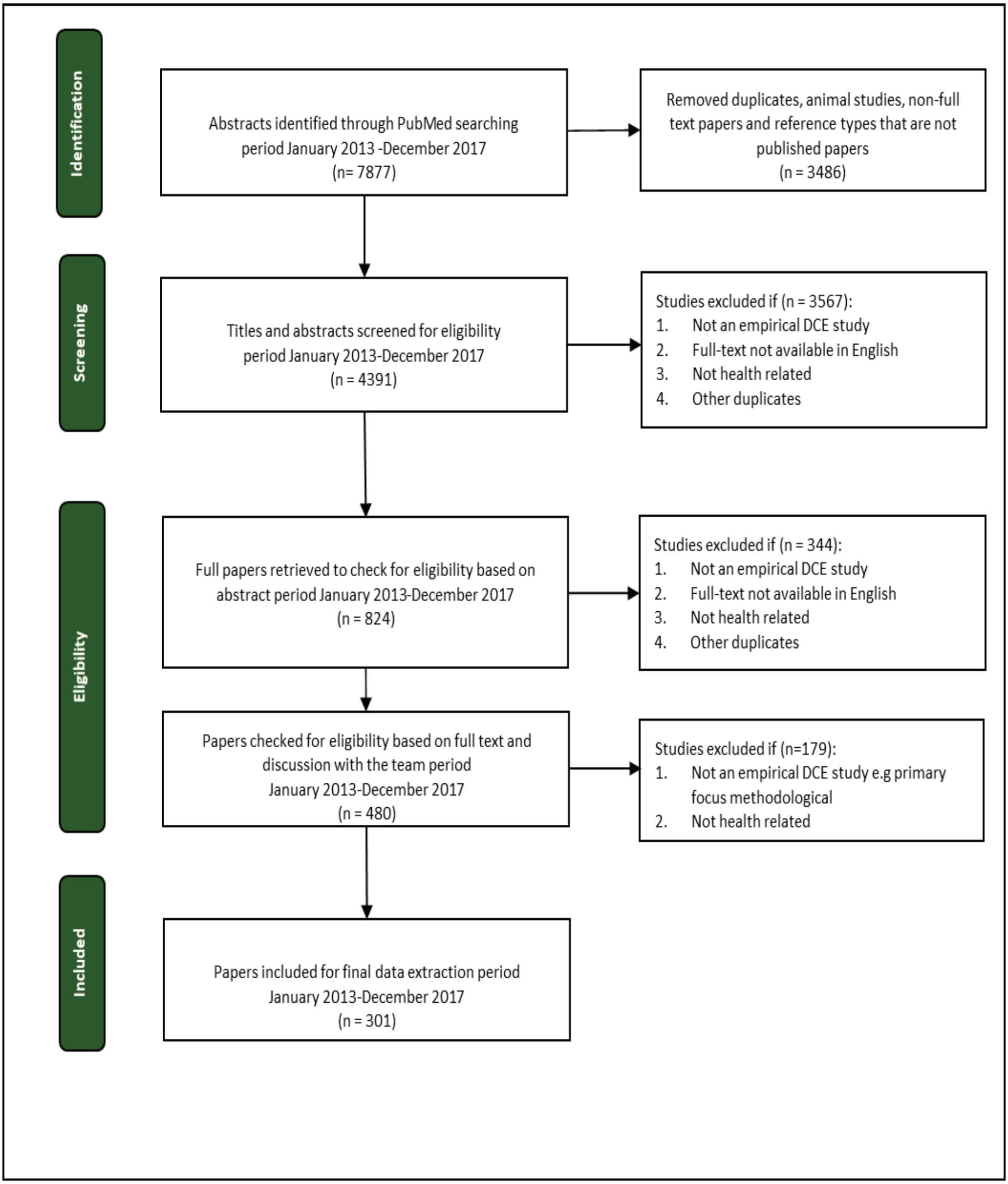


Fig. 2 Flow diagram of systematic literature review to identify DCEs

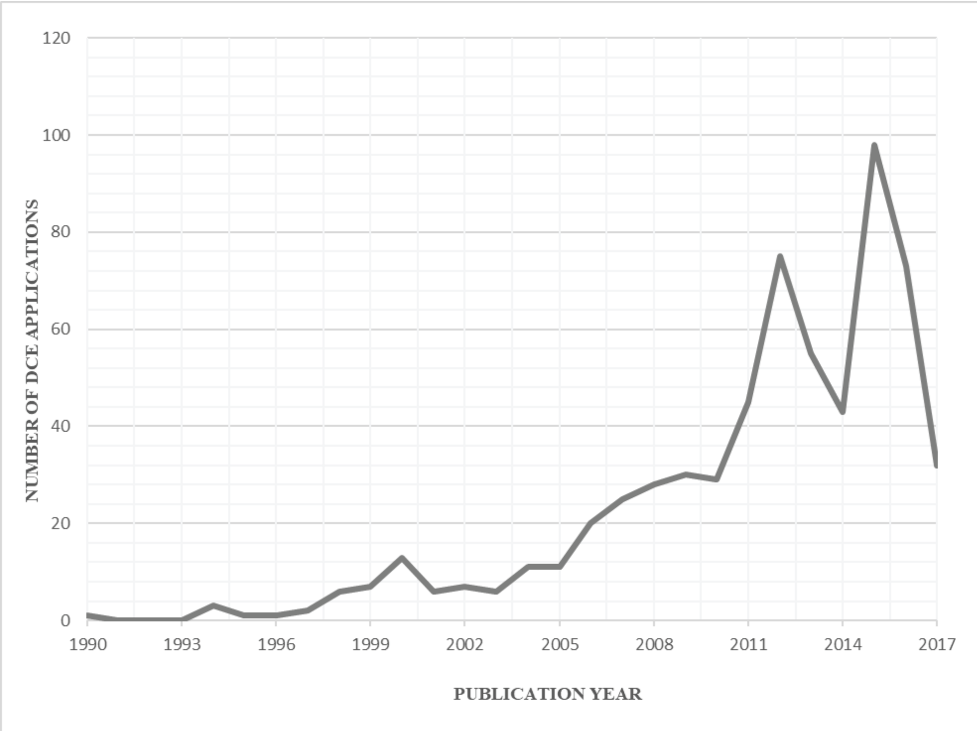


Fig. 3 Number of DCE applications by publication year

3.3.2 Areas of application

Prior reviews mentioned that although DCEs were originally introduced in health economics to value patient or consumer experience, the use of DCEs has broadened considerably [6,41]. Table 1 summarizes information about the different areas of application of DCEs for each review period (Appendix B of the Electronic Supplementary Material contains figures based on the tables in this review). Compared to the latest review period, the largest overall shifts occurred in the areas of patient consumer experience (category A), trade-offs between health outcomes and patient or consumer experience factors (category C), and health professionals' preferences for treatment or screening (category G). In the current review period, 8% of studies valued health outcomes such as 'heart attacks avoided' (category B, 23 studies, e.g. studies [169,173,174,183,191]), 4% estimated utility weights within the QALY framework (D, 13 studies, e.g. [239,247,248,249,251]), 6% focused on job choices (E, 17 studies, e.g.

[252,257,259,263,268]), and 9% developed priority-setting frameworks (F, 27 studies, e.g. [269,274,291,293,295]).

Among the DCEs reviewed, the most common journal focus was health services research (n=139; 46%). About a third (n=102; 34%) of articles were published in specialty-focused medical journals such as Vaccine (5 studies, [87,152,167,332,334]) or the British Journal of Cancer (3 studies, [68,91,192]). Fifty-one (17%) were published in general medical journals such as Plos One (20 studies, e.g. [65,85,102,112,120) and BMJ Open (5 studies, [121,123,130,190,285]). More details can be found in Appendix C of the Electronic Supplementary Material.

Table 1 Areas of study application

Category	Category details	Number of papers 1990-2000		Number of papers 2001-2008		Number of papers 2009-2012		Number of papers current: 2013-2017		Study numbers current: 2013-2017
		N=34[a]	*(%)*[b]	*N=114*[a]	*(%)*[b]	*N=179*[a]	*(%)*[b]	*N=301*[a]	*(%)*[b]	*N=301*[a]
A	Patient consumer experience	12	(35)	40	(35)	24	(12)	105	(35)	[64-168]
B	Valuing health outcomes	3	(9)	8	(7)	13	(7)	23	(8)	[169-191]
C	Investigating trade-offs between health outcomes and patient or consumer experience factors	14	(41)	38	(33)	81	(45)	47	(16)	[192-238]
D	Estimating utility weights within the QALY framework	0	(0)	2	(2)	4	(2)	13	(4)	[239-251]
E	Job choices	2	(6)	5	(4)	8	(4)	17	(6)	[252-268]
F	Developing priority setting frameworks	2	(6)	6	(5)	23	(13)	27	(9)	[269-295]
G	Health professionals' preferences for treatment or screening options for patients	1	(3)	17	(15)	23	(13)	19	(6)	[296-314]
H	Other	0	(0)	4	(4)	21	(12)	46	(15)	[315-360]
I	Not reported	N/C	N/C	N/C	N/C	N/C	N/C	4	(1)	[361-364]

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

3.3.3 Background information about DCEs

The reviews from Ryan & Gerard [11], de Bekker-Grob *et al.* [7] and Clark *et al.*[6] provided detailed information about study characteristics. Information for the current review period is described in the sections below. Tables 2a and 2b report the current information alongside data from the prior reviews.

Table 2a DCE Background information

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)*[b]	*N=114*[a]	*(%)*[b]	*N=179*[a]	*(%)*[b]	*N=301*[a]	*(%)*[b]
Country of origin	Australia	6	(18)	13	(11)	14	(8)	30	(10)
	Canada	1	(3)	6	(5)	23	(13)	25	(8)
	Germany	0	(0)	3	(3)	18	(10)	28	(9)
	Netherlands	0	(0)	5	(4)	27	(15)	44	(15)
	UK	20	(59)	55	(48)	39	(22)	50	(17)
	US	7	(21)	14	(12)	28	(16)	50	(17)
	Other	0	(0)	13	(11)	45	(25)	102	(34)
Number of attributes	2-3	5	(15)	15	(13)	14	(8)	30	(10)
	4-5	10	(29)	50	(44)	57	(32)	117	(39)
	6	9	(26)	30	(26)	61	(34)	67	(22)
	7-9	4	(12)	15	(13)	41	(23)	63	(21)
	10	2	(6)	2	(2)	5	(3)	4	(1)
	> 10	4	(12)	2	(2)	5	(3)	12	(4)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	8	(3)
Attributes covered	Monetary measure	19	(56)	61	(54)	102	(57)	150	(50)
	Time	25	(74)	58	(51)	118	(66)	117	(39)
	Risk	12	(35)	35	(31)	106	(59)	133	(44)
	Health status	19	(56)	62	(54)	109	(61)	71	(24)
	Health care	28	(82)	79	(69)	129	(72)	104	(35)
	Other	3	(9)	17	(15)	88	(49)	144	(48)
Number of choices per individual	8 or less	13	(38)	45	(39)	36	(20)	86	(29)
	9-16 choices	18	(53)	43	(38)	113	(63)	162	(54)
	> 16 choices	2	(6)	21	(18)	30	(17)	44	(15)
	Not clearly reported	1	(3)	5	(4)	5	(3)	9	(3)
Administration of survey	Self-completed questionnaire (paper)	27	(79)	76	(67)	86	(48)	69	(23)
	Self-completed questionnaire (online)	3	(9)	13	(11)	75	(42)	172	(57)
	Interviewer administered	3	(9)	22	(19)	34	(19)	44	(15)
	Other	N/C	N/C	N/C	N/C	N/C	N/C	5	(2)
	Not clearly reported	1	(3)	9	(8)	7	(4)	11	(4)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

Table 2b DCE Background information

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
Number of alternatives (not including opt-out/status quo)	2	N/C	N/C	N/C	N/C	N/C	N/C	251	(83)
	3	N/C	N/C	N/C	N/C	N/C	N/C	20	(7)
	4	N/C	N/C	N/C	N/C	N/C	N/C	5	(2)
	5	N/C	N/C	N/C	N/C	N/C	N/C	2	(1)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	23	(8)
Number of studies with opt-out/status quo	Yes	N/C	N/C	N/C	N/C	N/C	N/C	98	(33)
	No	N/C	N/C	N/C	N/C	N/C	N/C	194	(64)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	9	(3)
Sample size	Mean	N/C	N/C	N/C	N/C	N/C	N/C	728	N/A
	Median	N/C	N/C	N/C	N/C	N/C	N/C	401	N/A
Type of sample	Patients	15	(44)	N/C	N/C	N/C	N/C	110	(37)
	Healthcare workers	N/C	N/C	N/C	N/C	N/C	N/C	39	(13)
	General public	11	(32)	N/C	N/C	N/C	N/C	81	(27)
	Other	8	(24)	N/C	N/C	N/C	N/C	93	(31)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	5	(2)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category, N/A = Not applicable

3.3.3.1 Country of origin

Table 2a shows that UK-based studies made up a relatively high proportion of published DCEs (17%, 50 studies), as did the US (17%, 50 studies), Australia (10%, 30 studies), the Netherlands (15%, 44 studies), Germany (9%, 28 studies) and Canada (8%, 25 studies). DCEs were also popular in other European countries, for example Italy (3%, 8 studies) and Sweden (2%, 6 studies) (not shown). We also observed an increase in studies coming from 'other' countries, from 0% to 34% across the four review periods, which reflects an upwards trend towards applying DCEs in middle- and low-income countries (e.g. Cameroon [260], Ghana [265], Laos [253], Malawi [275] and Vietnam [143]).

3.3.3.2 Attributes, choices and survey

In the current review period the number of attributes per alternative in DCEs ranged from 2 to 21, with a median of 5. We observed a slight decrease in number of attributes; the modal category was 4-5 (39%, 117 studies). In line with prior reviews, most studies (82%, 247 studies) included 4 to 9 attributes. For the period 2013-2017, most studies included a monetary (50%, 150 studies), time-related (39%, 117 studies), or risk-related (44%, 133 studies) attribute. The proportion of studies including time-related and 'health status' (24%, 71 studies) attributes decreased.

Most DCEs in the current period included 9-16 choices per individual (54%, 162 studies), with a median of 12 (minimum:1, maximum:32). Prior reviews mentioned increases in the online administration of DCEs. This trend continued in the current review period, with 57% of the DCEs conducted online (172 studies), whereas the number of DCEs which used pencil and paper dropped to 23% (69 studies). These self-completed DCEs remained the main source of survey administration.

3.3.3.3 Alternatives and sample

Prior reviews did not collect data about the number of alternatives included in each DCE or whether an opt-out or status quo option was included. For the current period, most of the studies (83%, 251 studies) included 2 alternatives (not including any opt-out or status quo option), with 8% (23 studies) not clearly reporting the number of included alternatives (Table 2b). The majority of the studies (64%, 194 studies) did not include an opt-out or status quo option.

The prior reviews covering the periods 1990-2012 did not extract data about the sample size. In the current period, the mean and median sample size were 728 and 401 respectively. Sample size ranged from a minimum of 35 [137] to a maximum of 30,600 respondents [169]. Most of the samples included patients (37%, 110 studies) or the general public (27%, 81 studies). A large number of DCEs sampled 'other' populations (31%, 93 studies) such as healthcare workers, healthcare students or a mixture of these.

3.3.4 Experimental design

Experimental design (planning of the alternatives and choice sets) is crucial to the conduct of a DCE. The review from de Bekker-Grob *et al.* [7] describes DCE design in detail. For more information about the choices researchers have to make when designing the experimental part of a DCE, we also refer to a key checklist and best practice example [14,15].

3.3.4.1 Design type, design plan and blocking

As in prior review periods, most DCEs made use of a fractional design (89%, 269 studies) (Table 3). Additionally, we observed that for the current review period the design plan of DCEs most frequently focused on main effects only (29%, 86 studies). This is a decrease compared to the periods 1990-2000, 2001-2008 and 2009-2012, with 74%, 89% and 55% respectively. The percentage of DCEs not clearly reporting design plan information increased to 49% (147 studies) for 2013-2017. When generating the experimental design, blocking, creating different versions of the experiment for different respondent groups, can be used to reduce the cognitive burden of respondents by reducing the total number of choices per respondent [42]. Reviews for the period 1990-2012 did not collect information about blocking. Data for the current period showed that 50% (150 studies) reported using blocking when generating the experimental design. On average, studies with blocking had 709 participants, each of whom completed 11 choice sets, whereas studies with unblocked designs had 439 participants, each of whom completed 13 choice sets.

3.3.4.2 Design software

Ngene became the most popular software tool in the current period for generating experimental designs (21%, 62 studies: e.g. [74,84,160,289,340]). SAS (18%, 54 studies: e.g. [283,311,317,321,337]) and Sawtooth (16%, 47 studies: e.g. [67,162,228,297,344]) remained

Table 3 Experimental design information DCEs

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
Design type	Full factorial	4	(12)	0	(0)	9	(5)	13	(4)
	Fractional	25	(74)	114	(100)	158	(88)	269	(89)
	Not clearly reported	5	(15)	0	(0)	12	(7)	19	(6)
Design plan	Main effects only	25	(74)	100	(89)	98	(55)	86	(29)
	Main effects & two-way interactions	2	(6)	6	(5)	23	(13)	52	(17)
	Not applicable	4	(12)	0	(0)	5	(3)	5	(2)
	Other	N/C	N/C	N/C	N/C	N/C	N/C	11	(4)
	Not clearly reported	3	(9)	8	(7)	52	(29)	147	(49)
Blocking	Yes	N/C	N/C	N/C	N/C	N/C	N/C	150	(50)
	No	N/C	N/C	N/C	N/C	N/C	N/C	60	(20)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	91	(30)
Design software	Ngene	N/C	N/C	N/C	N/C	N/C	N/C	62	(21)
	SAS	0	(0)	14	(12)	41	(23)	54	(18)
	Sawtooth	2	(6)	5	(4)	30	(17)	47	(16)
	SPEED	13	(38)	22	(19)	9	(5)	1	(0)
	SPSS	2	(6)	14	(12)	13	(7)	20	(7)
	Not applicable	N/C	N/C	N/C	N/C	N/C	N/C	11	(3)
	Other	2	(6)	N/C	N/C	27	(15)	7	(2)
	Not clearly reported	N/C	N/C	4	(4)	9	(5)	99	(33)
Design source	Website	0	(0)	3	(3)	9	(5)	4	(1)
	Expert	4	(12)	4	(4)	11	(6)	5	(2)
	Not clearly reported	9	(26)	42	(37)	30	(17)	215	(71)
Methods to create choice sets	Orthogonal: Single profiles (binary choices)	3	(9)	12	(11)	2	(1)	7	(2)
	Orthogonal: Random pairing	18	(53)	19	(17)	18	(10)	12	(4)
	Orthogonal: Pairing with constant comparator	6	(18)	23	(20)	5	(3)	0	(0)
	Orthogonal: Foldover-random pairing	0	(0)	1	(1)	4	(2)	2	(1)
	Orthogonal: Foldover	0	(0)	11	(10)	34	(19)	26	(9)
	D-efficiency	0	(0)	14	(12)	54	(30)	105	(35)
	Bayesian D-efficiency	N/C	N/C	N/C	N/C	N/C	N/C	23	(8)
	Other	4	(12)	2	(2)	27	(15)	26	(9)
	Not clearly reported	3	(9)	32	(28)	39	(22)	100	(33)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

popular tools. Compared to prior review periods, we observed an increase in the percentage of studies not clearly indicating what software was used to generate the experimental design (33%, 99 studies: e.g. [65,165,198,225,320]).

3.3.4.3 Methods to create choice sets

The upwards trend in the use of D-efficient (35%, 105 studies) experimental designs continued in the current review period. Correspondingly, fewer DCEs used orthogonal arrays through methods such as single profiles, random pairing or the foldover technique (Table 3). As with the experimental design characteristics mentioned in the previous sections, we observed that an increasing number of studies (33%, 100 studies in 2013-2017) did not clearly report the methods used to create choice sets.

3.3.5 Econometric analysis methods

Information about the different econometric analysis methods and the appropriateness of these methods for different DCE applications is described in great detail in the prior reviews [6,7,11]. More information can be found in papers by Louviere & Lancsar [12], Bridges *et al.* [14] and Hauber *et al*. [17]. Tables 4a and 4b summarize information about econometric analyses from the current and prior review periods.

3.3.5.1 Econometric analysis model, software and preference heterogeneity

We present information about econometric analysis models according to the taxonomy described in the Methods section and visualized in Figure 1. Reviews for the period 1990-2000 and 2001-2008 reported that most DCEs used random-effects (random-intercept) probit models to analyse preference data (53% and 41% respectively). The review for the period 2009-2012 showed a shift to the use of other methods like multinomial logit models (45%) and mixed (random-parameter) logit models (25%). For the current review period, this trend continued (see Table 4a). Most DCEs in 2013-2017 reported the use of mixed logit models (39%, 118 studies: e.g. [68,292,322,335,339]) or multinomial logit models (39%, 116 studies: e.g. [113,131,187,315,360]) to analyse preference data. The current review period also showed an increase in the use of latent class models (12%, 36 studies: e.g. [40,112,160,186,290]) and other econometric analysis models. Examples include generalised multinomial logit (4%, 12 studies: e.g.: [118,145,178,195,261]) and heteroskedastic multinomial logit (4%, 11 studies: e.g. [155,160,205, 277,330]).

Prior reviews did not collect data about the software used for econometric analysis. For the current review period, Table 4b shows that most DCEs made use of Stata (31%, 94 studies: e.g. [112,131,159,170,234]) or Nlogit (22%, 65 studies: e.g. [115,192,225,303,56]) to conduct econometric analysis. However, 26% (79 studies: e.g. [122,205,252,232,351]) did not clearly report information about the software used.

Among the studies that used mixed logit models to account for preference heterogeneity in the period 2013-2017, 22% (65 studies) included additional information about the distributional assumptions used to conduct the mixed logit analysis and the number of distributional draws (e.g. Halton draws) used to simulate preference heterogeneity. This percentage is similar to the percentage for the period 2009-2012, which was 21%. The mean number of draws for the current review period was 1354 (median 1000, minimum 50, maximum 10,000) and 18% of the DCEs (53 studies) assumed that parameters followed the normal distribution.

3.3.6 Validity checks and qualitative methods

DCEs are based on responses to hypothetical choices (stated preferences), so internal and external validity checks provide a crucial opportunity to assess data quality or to compare stated preferences from DCEs with revealed preferences. As Clark *et al.* [6] observed in their review, there is often little reported about the tests for external validity possibly because validating hypothetical choice scenarios is difficult [44]. Perhaps for this reason, the review covering the period 1990-2000 did not extract specific information about external validity

Table 4a Econometric analysis details DCEs

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
Econometric analysis model	Random effects probit (random intercept)	18	(53)	47	(41)	18	(10)	17	(6)
	Logit	1	(3)	13	(11)	18	(10)	0	(0)
	Multinomial logit	6	(18)	25	(22)	86	(45)	116	(39)
	Random effects logit (random intercept)	1	(3)	6	(5)	14	(8)	15	(5)
	Mixed logit (random parameter)	1	(3)	6	(5)	45	(25)	118	(39)
	Latent class	0	(0)	1	(1)	7	(4)	36	(12)
	Nested logit	0	(0)	5	(4)	4	(2)	6	(2)
	Scale adjusted latent class	N/C	N/C	N/C	N/C	N/C	N/C	2	(1)
	Heteroskedastic multinomial logit	N/C	N/C	N/C	N/C	N/C	N/C	11	(4)
	Generalised multinomial logit	N/C	N/C	N/C	N/C	N/C	N/C	12	(4)
	Probit	6	(18)	8	(7)	4	(2)	7	(2)
	Other	1	(3)	4	(4)	32	(18)	25	(8)
	Not clearly reported	2	(6)	4	(4)	2	(1)	7	(2)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

Table 4b Econometric analysis details DCEs

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
Software for econometric analysis	Nlogit	N/C	N/C	N/C	N/C	N/C	N/C	65	(22)
	Biogeme	N/C	N/C	N/C	N/C	N/C	N/C	5	(2)
	Sawtooth	N/C	N/C	N/C	N/C	N/C	N/C	16	(5)
	R	N/C	N/C	N/C	N/C	N/C	N/C	10	(3)
	Stata	N/C	N/C	N/C	N/C	N/C	N/C	94	(31)
	SAS	N/C	N/C	N/C	N/C	N/C	N/C	17	(6)
	Other	N/C	N/C	N/C	N/C	N/C	N/C	15	(5)
	Not clearly reported	N/C	N/C	N/C	N/C	N/C	N/C	79	(26)
Mixed logit/random parameter logit-additional information	Number of studies with additional information	N/C	N/C	N/C	N/C	38	(21)	65	(22)
	Mean number of draws	N/C	N/C	N/C	N/C	N/C	N/C	1354	N/A
	Median number of draws	N/C	N/C	N/C	N/C	N/C	N/C	1000	N/A
	Distributional assumption: Normal distribution	N/C	N/C	N/C	N/C	20	(52)	53	(18)
	Distributional assumption: Other distribution/unclear	N/C	N/C	N/C	N/C	19	(50)	12	(4)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category, N/A = Not applicable

tests. In the reviews from 2001-2012 only a very small proportion (1%) of the DCEs reported any details about their investigations into external validity. The current review period showed that 2% (7 studies: [76,114,168,205,206,216,269]) reported using external validity tests (Table 5).

For detailed information about the different internal validity tests, we refer to the prior review papers [6,7,11]. In the current review period, the percentage of studies that included internal validity checks ranged from a maximum of 17% (50 studies) for non-satiation checks to 6% (18 studies) for internal compensatory checks. Internal compensatory checks were reported less frequently than in earlier review periods. For the current review period 'other' validity checks such as tests for theoretical and face validity and consistency were used frequently (34%, 102 studies).

Another way to enhance quality in a DCE is to complement the quantitative study with qualitative methods [37]. For the current review period, 86% (258) of the DCEs used qualitative methods to enhance the process and/or results. Most DCEs used interviews (50%, 151 studies) or focus group techniques (18%, 54 studies). Qualitative methods were usually used to inform attribute (53%, 160 studies) and/or level (44%, 134 studies) selection, which follows the overall upwards trend reported in prior reviews. The proportion of DCEs using qualitative methods for questionnaire pre-testing (38%, 113 studies) was similar to the level in the previous review period. Overall, just as in the previous review periods, few studies in the current review period (4%, 12 studies) used qualitative methods to improve the understanding of results/responses.

Table 5 Details of validity checks and qualitative methods

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
External validity tested	Yes	0	(0)	1	(1)	2	(1)	7	(2)
	No	34	(100)	113	(99)	177	(99)	294	(98)
Internal validity tested	Non-satiation (dominated questions)	15	(44)	56	(49)	36	(20)	50	(17)
	Transitivity (a>b, b>c then c>a)	3	(9)	5	(4)	2	(1)	2	(1)
	Sen's expansion and contraction	0	(0)	2	(2)	2	(1)	2	(1)
	Internal compensatory (1 attribute)	12	(35)	36	(32)	30	(17)	18	(6)
	Other	N/C	N/C	N/C	N/C	N/C	N/C	102	(34)
	Not clearly reported/not tested	N/C	N/C	N/C	N/C	N/C	N/C	189	(63)
Type of qualitative method used	Interviews	N/C	N/C	N/C	N/C	N/C	N/C	151	(50)
	Focus groups	N/C	N/C	N/C	N/C	N/C	N/C	54	(18)
	Other	N/C	N/C	N/C	N/C	N/C	N/C	53	(18)
	No qualitative method used	N/C	N/C	N/C	N/C	N/C	N/C	43	(14)
Rationale using qualitative methods	Attribute selection	6	(18)	79	(69)	90	(50)	160	(53)
	Level selection	6	(18)	38	(33)	73	(41)	134	(44)
	Pre-testing questionnaire	16	(47)	36	(32)	73	(41)	113	(38)
	Understanding results/responses	0	(0)	5	(4)	14	(8)	12	(4)
	Not clearly reported/other	N/C	N/C	N/C	N/C	N/C	N/C	5	(2)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

3.3.7 Outcome measures

Information about the trends regarding the presented outcome measures is presented in Table 6. As mentioned in prior reviews, DCEs often presented their outcomes in terms of willingness to pay (WTP), a monetary welfare measure or a utility score [6,7,11]. Use of these methods has declined over the past two review periods (2001-2012) and use of utility scores decreased from 24% to 8% over the past three periods (1990-2012). Relative to the previous period, we observed increases in the use of utility scores (17%, 50 studies, e.g. [82,149,162,185,338]), odds ratios (10%, 30 studies, e.g. [101,167,221,255,301]) and probability scores (13%, 38 studies: e.g. [143,175,219,293,298]). We also collected information about willingness to accept (WTA) measures (4%, 13 studies: e.g. [74,115,271,343,359]) and regression coefficients (56%, 169 studies: e.g.

[65,78,252,265,297]), which were not collected in previous reviews. The proportion of studies with 'other' outcome measures remained near one half (49%, 147 studies, e.g. [69,108,135,228,294]). Examples from this category include (predicted) choice shares, maximum acceptable risk, relative importance and ranking.

Table 6 Presented outcome measures of DCEs

Item	Category	1990-2000		2001-2008		2009-2012		Current: 2013-2017	
		N=34[a]	*(%)[b]*	*N=114[a]*	*(%)[b]*	*N=179[a]*	*(%)[b]*	*N=301[a]*	*(%)[b]*
Presented outcome measure	Per WTP unit	10	(29)	44	(39)	54	(30)	80	(27)
	Per WTA unit	N/C	N/C	N/C	N/C	N/C	N/C	13	(4)
	Per risk unit	3	(9)	2	(2)	4	(2)	9	(3)
	Monetary welfare measure	5	(15)	14	(12)	4	(2)	8	(3)
	Utility score	8	(24)	18	(16)	14	(8)	50	(17)
	Odds ratio	1	(3)	9	(8)	14	(8)	30	(10)
	Probability score	1	(3)	15	(13)	14	(8)	38	(13)
	Coefficients	N/C	N/C	N/C	N/C	N/C	N/C	169	(56)
	Other	N/C	N/C	N/C	N/C	90	(50)	147	(49)

[a]Numbers of individual studies might not add up to total Ns as some studies addressed multiple topics.
[b]Percentages might not add up to 100% because some studies addressed multiple topics and because of rounding error.
N/C = Not collected: data were not collected for this specific category

3.4 Discussion

In this study we reviewed DCEs published between 2013 and 2017. We followed the methods of prior reviews and compared our extraction results to those reviews to identify trends. We identified that DCEs have continued to increase in number and have been undertaken in more and more countries. Studies reported using more sophisticated designs with associated software, for example, D-efficient designs generated using Ngene. The trend towards the use of more sophisticated econometric models has also continued. However, many studies presented sophisticated methods with insufficient detail. For example, we were not able to check whether the results had the correct interpretation or whether the authors had conducted the appropriate diagnostics (e.g. checked that the data possessed the IIA characteristic). Qualitative methods have continued to be popular as an approach to select attributes and levels, which might improve validity. In this study we also extracted data in several new categories, for example, sample size and type, the use of blocking, software used for econometric analysis and type of qualitative method used. We observed that the mean and median sample size were 728 and 401 respectively, with most samples including patients. We also observed that half of the studies used blocking and most studies used Stata for econometric analysis. Interviewing was the most popular qualitative research method used alongside DCEs.

The observed increase in the total number of DCEs in health economics was similar to the trend reported in prior reviews [6,7,11], but less consistent from year to year (Figure 3). This less consistent increase might be explained by the presence of many competing stated preference methods [4,5,44]. We hypothesize that other methods may be increasing in popularity or becoming more useful in health settings [45]. Examples of such methods may include best-worst-scaling (BWS) case 1 and case 2 [46-48], which were not included in this review. Additionally, in this review we excluded a significant number (31) of studies making methodological considerations about DCEs rather than conducting empirical research. The presence of such studies may indicate that knowledge about DCEs in health has increased and there is more focus on studies to develop the method. Examples include simulation studies about experimental design, studies comparing the outcomes of a DCE to other stated preference method outcomes and studies examining different model specifications [49-51]. This might be another explanation for the less consistent increase in DCE application studies.

The common use of fractional designs, as described in prior reviews [6,7], has continued. This review also found that main effects DCEs continue to dominate; however, there is a downwards trend as DCE designs incorporate two-way interactions more often. This is in line with the recommendations of Louviere and Lancsar [12] who suggest the inclusion of interaction terms should be explored in the experimental design stage. Ngene became the most popular software tool in the current review period for generating experimental designs, while D-efficient designs became the most popular method to create choice sets. Perhaps as a consequence of the rise in software-generated designs, this review also showed that an increasing percentage of articles did not include information about experimental design features such as the design plan. Omitting this type of information might inhibit quality assessment and reduce confidence in the results. Future research might focus on the specific reasons why such information is missing and the impact of the missing information on quality assessment of DCEs. One potential reason for omitting methodological details is the journal word limit. When confronted with a low word limit, authors should consider using online space to report additional design and analysis details.

In addition to these observations about the generation of experimental designs, we identified design information that would be helpful to report in DCEs and future systematic reviews. For example, prior reviews did not include information about blocking, and although at least half of the DCEs we reviewed used blocking, 30% of the studies we reviewed did not include information about blocking. Blocking could be an important technique in light of the growing literature about the cognitive burden of DCEs and the impact of this cognitive burden on respondent outcomes [42]. However, blocking also has the disadvantage of requiring a larger sample size [42]. The approach described by Sandor & Wedel [52] might be another alternative to increase the validity of DCE outcomes in case of relatively small sample sizes or the investigation of preference heterogeneity.

Prior reviews identified a shift to more flexible econometric analysis models [6,7], which is not necessarily positive. This trend has continued in this review. Most studies included multinomial logit or mixed logit models. Although we did not formally extract information about variance estimation, we noted that among the DCEs using multinomial logit models to analyse choice data, few reported robust or Huber-White standard errors (most studies reported ‘regular’ standard errors). Since these standard errors allow for more flexible substitution patterns and flexible variances, it is common in economics and econometrics to report these standard errors instead of ‘regular’ standard errors [53]. Also, in the presence of repeated observations from the same individuals, conventional standard errors are biased downward [54]. Thus, future DCEs in health economics could benefit from more appropriate treatment of clustered data (i.e., use of robust standard errors) and more complete reporting of econometric output.

In terms of analytical methods, we also observed some patterns in the exploration of preference and scale heterogeneity. We noted that, among the 39% of studies that used a mixed logit model, many treated heterogeneity as a nuisance, i.e., they used the mixed model to accommodate repeated measures but did not report additional information about the "mixed" aspect of the data (e.g. standard deviation estimates). Since preference heterogeneity is regarded as an important aspect within choice modelling, taking full advantage of the modelling results might help us understand preference heterogeneity better [55]. With regard to scale heterogeneity, work by Fiebig *et al*. [56] indicated that other models such as the generalised multinomial logit and heteroskedastic multinomial logit models could be considered when analysing DCE data, to identify differences in scale when comparing preferences between groups of respondents [57]. Data from this review identified a small number of DCEs using such methods; for a more detailed breakdown we refer readers to another review focussing on scale heterogeneity specifically [32]. However, it is important to mention that the generalised multinomial logit model should be used with caution since the ability of this model to capture scale heterogeneity has been questioned in the literature [58].

Articles by Vass & Payne [20] and Mott [21] describe issues influencing the degree to which DCE findings are used in healthcare decision-making (e.g. health-state valuation and health technology assessment). These articles, rising popularity of the method, and interest from regulators and funders suggest that DCEs could play an important role in real-world decision-making [59,60]. However, concerns have been expressed about the validity, reliability, robustness and generalisability of DCEs [11,61]. A key stage in understanding the robustness of DCEs is understanding whether stated preferences reflect ‘true’ preferences as revealed in the market [10]. In this study we observed that the number of studies testing external validity remained small. Future research should focus on identifying and resolving the methodological and practical challenges involved in validity testing, and on guiding the incorporation of DCEs into actual decision-making in healthcare. Another practice that may improve the robustness of DCEs and facilitate their use in healthcare decision-making, is the increased use of qualitative methods to complement quantitative DCE analysis [61]. Prior reviews and additional literature suggest that qualitative research methods can strengthen DCEs and other quantitative methods by facilitating numerous investigations such as 1) identification of relevant attributes and levels, 2) verification that respondents understand the presented information, and 3) learning about respondents’ decision strategies [6,7,11,62]. These investigations can help determine whether respondents are making choices in line with the underpinning utility theories, thereby supporting the legitimacy of the underlying assumptions. This review showed an overall upwards trend in the number of DCEs using qualitative methods to select attributes and levels. This move towards a more mixed-methods approach has been observed by others, for example the study by Ikenwilo *et al.* [63].

3.4.1 Strengths and limitations

The current study has several strengths. First, the detailed data extraction was completed by each author individually, with the total number of articles approximately divided equally among authors because of the relatively short timeframe and the need to balance author burden with study quality. Additionally, a subsample of studies (20%) was double-checked by one author (VS) for quality control, which enhanced reliability. Second, this study identified trends in empirical DCEs by comparing outcomes from all prior reviews. Additionally, this study included aspects of empirical DCEs not investigated before, although these aspects were recognized in the literature as becoming more important in DCE research (e.g. blocking in experimental design and the type of qualitative methods used in a DCE). Third, our observation of less rapid growth in the number of empirical DCEs (compared to the growth observed in previous reviews) matches the trend in the preference research to focus on the broad range of stated preference methods available (rather than DCEs exclusively) [4,5,44].

A potential weakness of this study was the use of multiple reviewers with potentially different interpretations of DCE reports which might have affected the data extraction and, as a consequence, the results presented. To limit inconsistency between reviewers, all co-authors discussed the data extraction frequently and results were cross-validated by a single author (VS). Similarly, this inconsistency in interpretation may also have occurred between the different review periods. Procedural information from the two most recent reviews was used to ensure consistency, and we are therefore confident the general trends reported and conclusion that more detailed methods reporting is called for holds. Another potential weakness is the use of only one database (PubMed). However, like the authors of the prior reviews [6,7], we do not expect the review findings to be significantly different when performing searches on other databases. Also, since we were interested in identifying trends and therefore maximising comparability between the different reviews, we preferred to restrict our searches to this single database. As with many systematic reviews, data were extracted from published manuscripts and online appendices. The results are therefore reliant on what was reported in the final article and do not necessarily reflect all activities of the authors. Trends presented could therefore reflect factors such as publication bias, journal scope, editor preferences, and word limits, as well as preferences of journal editors rather than actual practice. Additionally, although we did update the data extraction tool based on changes in the field, future research might benefit from updating other aspects of the systematic review protocol such as search terms and inclusion and exclusion criteria (e.g. inclusion of best-best scaling). Finally, although we believe that DCEs are both useful and common enough to deserve focused attention in this review, DCEs represent one method among many for examining health preferences, and other methods may be preferable depending on the circumstances [4].

3.5 Conclusion

This study provides an overview of the applications and methods used by DCEs in health. The use of empirical DCEs in health economics has continued to grow, as have the areas of application and the geographic scope. This study identified changes in the experimental design (e.g. more frequent use of D-efficient designs), analysis methods (e.g. mixed logit models most frequently used), validity enhancement (e.g. more diverse use of internal validity checks), qualitative methods (e.g. upwards trend of qualitative methods used for attribute and level selection) and outcome measures (e.g. coefficients most frequently used). However, a large number of studies not reporting methodological details were also identified. DCEs should include more complete information, for example information about design generation, blocking, model specification, random-parameter estimation and model results. Developing reporting guidelines specifically for DCEs might positively impact quality assessment, increase confidence in the results and improve the ability of decision-makers to act on the results. How and when to integrate health-related DCE outcomes into decision-making remains an important area for future research.

References

1. Hogg CN. Patient and public involvement: what next for the NHS?. Health Expect. 2007 Jun;10(2):129-38.

2. Mockford C, Staniszewska S, Griffiths F, Herron-Marx S. The impact of patient and public involvement on UK NHS health care: a systematic review. International Journal for Quality in Health Care. 2012; Volume 24, Issue 1, Pages 28–38, https://doi.org/10.1093/intqhc/mzr066.

3. Soekhai V, Whichello C, Levitan B, Veldwijk J, Hammad T, Kihlbom U et al. Compendium Of Methods For Measuring Patient Preferences In Medical Treatment. Value in Health. 2017; Volume 20, Issue 9, Pages A684–A685.

4. MDIC. Medical Device Innovation Consortium (MDIC): Patient Centered Risk-Benefit Project Report. [Internet]. 2015. Available from: http://mdic.org/wpcontent/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf

5. Mahieu PA, Andersson H, Beaumais O, Crastes dit Sourd R, Hess S, Wolff, F. Stated preferences: a unique database composed of 1,657 recent published articles in journals related to agriculture, environment or health. Review of Agricultural, Food and Environmental Studies. 2017; 98 (3). pp. 201-220.

6. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. Pharmacoeconomics [Internet]. 2014 [cited 2018 Jun 18];32:883–902. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25005924

7. De Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: A review of the literature. Health Econ [Internet]. Wiley-Blackwell; 2012 [cited 2018 Jun 18];21:145–72. Available from: http://doi.wiley.com/10.1002/hec.1697.

8. Lancaster KJ. A New Approach to Consumer Theory. J Polit Econ [Internet]. The University of Chicago Press ; 1966 [cited 2018 Jun 18];74:132–57. Available from: http://www.journals.uchicago.edu/doi/10.1086/259131.

9. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: A systematic review. BMC Health Serv Res [Internet]. 2014 [cited 2018 Jun 18];14. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4161911/pdf/12913_2014_Article_3466.pdf

10. Thurstone LL. A law of comparative judgment. Psychol Rev [Internet]. 1927 [cited 2018 Jul 8];34:273–86. Available from: http://content.apa.org/journals/rev/34/4/273

11. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Heal Econ Heal Policy [Internet]. 2003 [cited 2018 Jun 18];2:55–64. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=14619274%5Cnhttp://sfxeu08.hosted.exlibrisgroup.com/sgul?sid=OVID:medline&id=pmid:14619274&id=doi:&issn=1175-5652&isbn=&volume=2&issue=1&spage=55&pages=55-64&date=2003&title=Appl

12. Louviere JJ, Lancsar E. Choice experiments in health: The good, the bad, the ugly and toward a brighter future. Heal Econ Policy Law [Internet]. 2009 [cited 2018 Jun 20];4:527–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19715635

13. Lancsar E, Swait J. Reconceptualising the External Validity of Discrete Choice Experiments. Pharmacoeconomics [Internet]. Springer International Publishing; 2014 [cited 2018 Jun 26];32:951–65. Available from: http://link.springer.com/10.1007/s40273-014-0181-7

14. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health - A checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.

Value Heal [Internet]. 2011 [cited 2018 Jun 23];14:403–13. Available from: https://www.ispor.org/TaskForces/documents/ISPOR-CA-in-Health-TF-Report-Checklist.pdf

15. Johnson FR, 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 Heal [Internet]. 2013 [cited 2018 Jun 23];16:3–13. Available from: https://www.ispor.org/ValueInHealth/ShowValueInHealth.aspx?issue=3551E3F3-83A8-4895-9415-5272AF29A9DF

16. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Health Care Decision-making : A User ' s Guide , Experiments to Inform Healthcare A User ' s Guide. Pharmacoeconomics [Internet]. 2008 [cited 2018 Jul 13];26:661–77. Available from: https://link.springer.com/content/pdf/10.2165%2F00019053-200826080-00004.pdf

17. Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Heal [Internet]. 2016 [cited 2018 Jun 23];19:300–15. Available from: https://www.ispor.org/Conjoint-Analysis-Statistical-Methods-Guidelines.pdf

18. Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? J Consum Res [Internet]. The University of Chicago Press; 2008 [cited 2018 Jun 20];35:360–75. Available from: https://academic.oup.com/jcr/article-lookup/doi/10.1086/586913.

19. Louviere JJ, Lancsar E. Choice experiments in health: The good, the bad, the ugly and toward a brighter future. Heal Econ Policy Law [Internet]. 2009 [cited 2018 Jun 20];4:527–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19715635

20. Vass CM, Payne K. Using Discrete Choice Experiments to Inform the Benefit-Risk Assessment of Medicines: Are We Ready Yet? Pharmacoeconomics [Internet]. 2017 [cited 2018 Jul 8];35:859–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28536955

21. Mott DJ. Incorporating Quantitative Patient Preference Data into Healthcare Decision-making Processes: Is HTA Falling Behind? Patient [Internet]. Springer International Publishing; 2018 [cited 2018 Jul 8];11:249–52. Available from: http://link.springer.com/10.1007/s40271-018-0305-9

22. Mulhern B, Norman R, Street D, Viney R. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. PharmacoEconomics. 2018;1-15.

23. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. The European journal of health economics. 2018;1-14.

24. Zhou M, Thayer WM, Bridges JF. Using latent class analysis to model preference heterogeneity in health: a systematic review. PharmacoEconomics. 2018;1-1.

25. Marshall D, Bridges JF, Hauber B, Cameron R, Donnalley L, Fyie K. et al. Conjoint analysis applications in health—how are studies being designed and reported?. The Patient: Patient-Centered Outcomes Research. 2010;3(4), 249-256.

26. Kleij KS, Tangermann U, Amelung VE, Krauth C. Patients' preferences for primary health care–a systematic literature review of discrete choice experiments. BMC health services research. 2017;17(1), 476.

27. Clark MD, Szczepura A, Gumber A, Howard K, Moro D, Morton RL. Measuring trade-offs in nephrology: a systematic review of discrete choice experiments and conjoint analysis studies. Nephrology Dialysis Transplantation. 2017 33(2), 348-355.

28. Michaels-Igbokwe C, MacDonald S, Currie GR. Individual preferences for child and adolescent vaccine attributes: A systematic review of the stated preference literature. The Patient-Patient-Centered Outcomes Research. 2017;10(6), 687-700.

29. Regmi, K, Kaphle D, Timilsina S, & Tuha, NAA. Application of discrete-choice experiment methods in tobacco control: a systematic review. PharmacoEconomics-open. 2018;1-13.

30. 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. 2017;1-17.

31. Harrison M, Milbers K, Hudson M, Bansback N. Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. BMJ open. 2017;7(5), e014719.

32. Wright SJ, Vass CM, Sim G, Burton M, Fiebig DG, Payne K. Accounting for Scale Heterogeneity in Healthcare-Related Discrete Choice Experiments when Comparing Stated Preferences: A Systematic Review. The Patient-Patient-Centered Outcomes Research. 2018; 1-14.

33. Bien DR, Danner M, Vennedey V, Civello D, Evers SM, Hiligsmann M. Patients' preferences for outcome, process and cost attributes in cancer treatment: A systematic review of discrete choice experiments. The Patient-Patient-Centered Outcomes Research. 2017;10(5), 553-565.

34. Vass C, Gray E, Payne K. Discrete choice experiments of pharmacy services: a systematic review. International journal of clinical pharmacy. 2016;38(3), 620-630.

35. Mansfield C, Tangka FK, Ekwueme DU, Smith JL, Guy Jr, GP, Li C et al. Peer Reviewed: Stated Preference for Cancer Screening: A Systematic Review of the Literature, 1990–2013. Preventing chronic disease. 2016;13.

36. Schatz, NK, Fabiano GA, Cunningham CE, Waschbusch DA, Jerome S., Lupas K, Morris KL. Systematic review of patients' and parents' preferences for ADHD treatment options and processes of care. The Patient-Patient-Centered Outcomes Research. 2015;8(6), 483-497.

37. Vass C, Rigby D, Payne K. The role of qualitative research methods in discrete choice experiments: a systematic review and survey of authors. Medical Decision-making. 2017;37(3), 298-313.

38. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. BMC health services research. 2014;14(1), 367.

39. Hiligsmann M, Bours SP, Boonen A. A review of patient preferences for osteoporosis drug treatment. Current rheumatology reports. 2015;17(9), 61.

40. Webb EJ, Meads D, Eskyte I, King N, Dracup N, Chataway J et al. A Systematic Review of Discrete-Choice Experiments and Conjoint Analysis Studies in People with Multiple Sclerosis. The Patient-Patient-Centered Outcomes Research. 2018:1-12.

41. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: An application to in vitro fertilization. Soc Sci Med [Internet]. 1999 [cited 2018 Jun 23];48:535–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10075178

42. Jonker MF, Donkers B, De Bekker-Grob EW, Stolk EA. The Effect of Level Overlap and Color Coding on Attribute Non-attendance in Discrete Choice Experiments. 2017 [cited 2018 Jun 27]; Available from: https://www.valueinhealthjournal.com/article/S1098-3015(17)33609-4/pdf

43. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res [Internet]. 2017 [cited 2018 Jul 25];17:531–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29058478

44. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, et al. Eliciting Public Preferences for Healthcare. Health Technol Assess (Rockv) [Internet]. 2001;5:1–186. Available from: http://www.hta.ac.uk/execsumm/summ505.htm

45. Cheung KL, Wijnen BFM, Hollin IL et al. Using Best–Worst Scaling to Investigate Preferences in Health Care. PharmacoEconomics. 2016;34: 1195.

46. Krucien N, Watson V, Ryan M. Is Best-Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments. Health Econ [Internet]. Wiley-Blackwell; 2017 [cited 2018 Jun 26];26:e1–16. Available from: http://doi.wiley.com/10.1002/hec.3459

47. Kremer IEH, Evers SMAA, Jongen PJ, Hiligsmann M. Comparison of preferences of healthcare professionals and MS patients for attributes of disease-modifying drugs: A best-worst scaling. Heal Expect [Internet]. Wiley/Blackwell (10.1111); 2018 [cited 2018 Jun 26];21:171–80. Available from: http://doi.wiley.com/10.1111/hex.12599

48. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Economics Review. 2016;6:2. doi:10.1186/s13561-015-0079-x.

49. Vanniyasingam T, Cunningham CE, Foster G, Thabane L. Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. BMJ Open [Internet]. British Medical Journal Publishing Group; 2016 [cited 2018 Jun 26];6:e011985. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27436671

50. Pignone MP, Howard K, Brenner AT, Crutchfield TM, Hawley ST, Lewis CL, et al. Comparing 3 Techniques for Eliciting Patient Values for Decision-making About Prostate-Specific Antigen Screening. JAMA Intern Med [Internet]. 2013 [cited 2018 Jul 16];173:362. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23400279

51. Lancsar E, Fiebig DG, Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. Pharmacoeconomics [Internet]. Springer-Verlag London Ltd.; 2017 [cited 2018 Jul 16];35:697–716. Available from: http://link.springer.com/10.1007/s40273-017-0506-4

52. Sándor Z, Wedel M. Designing Conjoint Choice Experiments Using Managers' Prior Beliefs. J Mark Res [Internet]. 2001 [cited 2018 Jul 30];38:430–44. Available from: http://journals.ama.org/doi/abs/10.1509/jmkr.38.4.430.18904

53. Pforr K. femlogit—Implementation of the multinomial logit model with fixed effect. Stata J [Internet]. 2014 [cited 2018 Jul 9];14:847–62. Available from: http://ideas.repec.org/a/tsj/stataj/v7y2007i4p465-506.html

54. Snijders TAB, Borgatti SP. Non-parametric standard errors and tests for network statistics. Connections. 1999;2:61 - 70.

55. Groothuis-Oudshoorn CGM, Flynn TN, Yoo H Il, Magidson J, Oppe M. Key Issues and Potential Solutions for Understanding Healthcare Preference Heterogeneity Free from Patient-Level Scale Confounds. Patient - Patient-Centered Outcomes Res [Internet]. Springer International Publishing; 2018 [cited 2018 Jul 9];1–4. Available from: http://link.springer.com/10.1007/s40271-018-0309-5

56. Fiebig DG, Keane MP, Louviere J, Wasi N. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Mark Sci [Internet]. INFORMS ; 2010 [cited 2018 Jul 9];29:393–421. Available from: http://pubsonline.informs.org/doi/abs/10.1287/mksc.1090.0508

57. Vass CM, Wright S, Burton M, Payne K. Scale Heterogeneity in Healthcare Discrete Choice Experiments: A Primer. Patient - Patient-Centered Outcomes Res [Internet]. Springer International Publishing; 2018 [cited 2018 Jul 16];11:167–73. Available from: http://link.springer.com/10.1007/s40271-017-0282-4

58. Hess S, Rose JM. Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation (Amst) [Internet]. Springer US; 2012 [cited 2018 Jul 30];39:1225–39. Available from: http://link.springer.com/10.1007/s11116-012-9394-9

59. de Bekker-Grob EW, Berlin C, Levitan B, Raza K, Christoforidi K, Cleemput I, et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. Patient - Patient-Centered Outcomes Res [Internet]. Springer International Publishing; 2017 [cited 2018 Jul 25];10:263–6. Available from: http://link.springer.com/10.1007/s40271-017-0222-3

60. Food and Drug Administration (FDA). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling [Internet]. 2016. Available from: http://www.regulations.gov.

61. Gonzalez JM, Johnson FR, Levitan B, Noel R, Peay H. Symposium Title: Preference Evidence for Regulatory Decisions. Patient. 2018 Oct;11(5):467-473. doi: 10.1007/s40271-018-0311-y.

62. Sandelowski, M. Combining Qualitative and Quantitative Sampling,Data Collection, and Analysis Techniques in Mixed-Method Studies. Research in Nursing & Health. 2000;23, 246–255.

63. Ikenwilo D, Heidenreich S, Ryan M, Mankowski C, Nazir J, Watson V. The Best of Both Worlds: An Example Mixed Methods Approach to Understand Men's Preferences for the Treatment of Lower Urinary Tract Symptoms. Patient Patient-Centered Outcomes Res [Internet]. 2018 [cited 2018 Jul 9];11:55–67. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=prem&AN=28660567 http://nt2yt7px7u.search.serialssolutions.com/?sid=OVID:Ovid+MEDLINE%28R%29+In-Process+%26+Other+Non-Indexed+Citations+%3CApril+27%2C+2018%3E&genre=article&id=pmid:286605

64. Augustovski F, Beratarrechea A, Irazola V, Rubinstein F, Tesolin P, Gonzalez J, et al. Patient preferences for biologic agents in rheumatoid arthritis: A discrete-choice experiment. Value Heal. 2013;16:385–93.

65. Boone D, Mallett S, Zhu S, Yao GL, Bell N, Ghanouni A, 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.

66. Carroll FE, Al-Janabi H, Flynn T, Montgomery AA. Women and their partners' preferences for Down's syndrome screening tests: A discrete choice experiment. Prenat Diagn. 2013;33:449–56.

67. Deal K, Marshall D, Dabrowski D, Charter A, Bukoski M, Moayyedi P. Assessing the value of symptom relief for patients with gastroesophageal reflux disease treatment: Willingness to pay using a discrete choice experiment. Value Heal [Internet]. Elsevier; 2013;16:588–98. Available from: http://dx.doi.org/10.1016/j.jval.2013.01.007

68. De Bekker-Grob EW, Bliemer MCJ, Donkers B, Essink-Bot ML, Korfage IJ, Roobol MJ, et al. Patients' and urologists' preferences for prostate cancer treatment: A discrete choice experiment. Br J Cancer [Internet]. Nature Publishing Group; 2013;109:633–40. Available from: http://dx.doi.org/10.1038/bjc.2013.370

69. Gelhorn HL, Stringer SM, Brooks A, Thompson C, Monz BU, Boye KS, et al. Preferences for medication attributes among patients with type 2 diabetes mellitus in the UK. Diabetes, Obes Metab. 2013;15:802–9.

70. Glenngård AH, Hjelmgren J, Thomsen PH, Tvedten T. Patient preferences and willingness-to-pay for ADHD treatment with stimulants using discrete choice experiment (DCE) in Sweden, Denmark and

Norway. Nord J Psychiatry. 2013;67:351–9.

71. Gonzalez JM, Johnson FR, Runken MC, Poulos CM. Evaluating migraineurs’ preferences for migraine treatment outcomes using a choice experiment. Headache. 2013;53:1635–50.

72. Hauber AB, Arden NK, Mohamed AF, Johnson FR, Peloso PM, Watson DJ, et al. A discrete-choice experiment of United Kingdom patients’ willingness to risk adverse events for improved function and pain control in osteoarthritis. Osteoarthr Cartil [Internet]. Elsevier Ltd; 2013;21:289–97. Available from: http://dx.doi.org/10.1016/j.joca.2012.11.007

73. Krucien N, Gafni A, Fleury B, Pelletier-Fleury N. Patients’ with obstructive sleep apnoea syndrome (OSAS) preferences and demand for treatment: A discrete choice experiment. Thorax. 2013;68:487–8.

74. Laba TL, Brien JA, Fransen M, Jan S. Patient preferences for adherence to treatment for osteoarthritis: The MEdication Decisions in Osteoarthritis Study (MEDOS). BMC Musculoskelet Disord. 2013;14.

75. Lathia N, Isogai PK, Walker SE, De Angelis C, Cheung MC, Hoch JS, et al. Eliciting patients’ preferences for outpatient treatment of febrile neutropenia: A discrete choice experiment. Support Care Cancer. 2013;21:245–51.

76. Meghani SH, Chittams J, Hanlon AL, Curry J. Measuring preferences for analgesic treatment for cancer pain: How do African-Americans and Whites perform on choice-based conjoint (CBC) analysis experiments? BMC Med Inform Decis Mak [Internet]. BMC Medical Informatics and Decision-making; 2013;13:1. Available from: BMC Medical Informatics and Decision-making

77. Milte R, Ratcliffe J, Miller M, Whitehead C, Cameron ID, Crotty M. What are frail older people prepared to endure to achieve improved mobility following hip fracture? a discrete choice experiment. J Rehabil Med. 2013;45:81–6.

78. Mohamed AF, Zhang J, Johnson FR, Lomon ID, Malvolti E, Townsend R, et al. Avoidance of weight gain is important for oral type 2 diabetes treatments in Sweden and Germany: Patient preferences. Diabetes Metab [Internet]. Elsevier Masson SAS; 2013;39:397–403. Available from: http://dx.doi.org/10.1016/j.diabet.2013.06.001

79. Moia M, Mantovani LG, Carpenedo M, Scalone L, Monzini MS, Cesana G, et al. Patient preferences and willingness to pay for different options of anticoagulant therapy. Intern Emerg Med. 2013;8:237–43.

80. Najafzadeh M, Johnston KM, Peacock SJ, Connors JM, Marra MA, Lynd LD, et al. Genomic testing to determine drug response: Measuring preferences of the public and patients using Discrete Choice Experiment (DCE). BMC Health Serv Res. 2013;13.

81. Parkinson B, Goodall S, Norman R. Measuring the loss of consumer choice in mandatory health programmes using discrete choice experiments. Appl Health Econ Health Policy. 2013;11:139–50.

82. Pisa G, Freytag S, Schandry R. Chronic obstructive pulmonary disease (COPD) patients’ disease-related preferences: A study using conjoint analysis. Patient. 2013;6:93–101.

83. Schaarschmidt ML, Umar N, Schmieder A, Terris DD, Goebeler M, Goerdt S, et al. Patient preferences for psoriasis treatments: Impact of treatment experience. J Eur Acad Dermatology Venereol. 2013;27:187–98.

84. Veldwijk J, Lambooij M, van Gils P, Struijs J, Smit H, de Wit GA. Type 2 diabetes patients’ preferences and willingness to pay for lifestyle programs: a discrete choice experiment. BMC Public Health [Internet]. 2013;13:1099. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=24289831

85. Whitty JA, Stewart S, Carrington MJ, Calderone A, Marwick T, Horowitz JD, et al. Patient Preferences and Willingness-To-Pay for a Home or Clinic Based Program of Chronic Heart Failure Management: Findings from the Which? Trial. PLoS One. 2013;8:1–8.

86. Zimmermann TM, Clouth J, Elosge M, Heurich M, Schneider E, Wilhelm S, et al. Patient preferences for outcomes of depression treatment in Germany: A choice-based conjoint analysis study. J Affect Disord [Internet]. Elsevier; 2013;148:210–9. Available from: http://dx.doi.org/10.1016/j.jad.2012.11.062

87. Cameron MP, Newman PA, Roungprakhon S, Scarpa R. The marginal willingness-to-pay for attributes of a hypothetical HIV vaccine. Vaccine [Internet]. Elsevier Ltd; 2013;31:3712–7. Available from: http://dx.doi.org/10.1016/j.vaccine.2013.05.089

88. Tokes K, Quadri S, Cahill P, Chiu G, Ivanov A, Tang H. Disease and treatment perceptions among Asian Americans diagnosed with chronic hepatitis B infection. J Gen Intern Med. 2014;29:477–84.

89. Dixon S, Nancarrow SA, Enderby PM, Moran AM, Parker SG. Assessing patient preferences for the delivery of different community-based models of care using a discrete choice experiment. 2013;1204–14.

90. Benning TM, Dellaert BGC, Severens JL, Dirksen CD. The effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: Results from a discrete choice experiment. Value Heal [Internet]. Elsevier; 2014;17:578–87. Available from: http://dx.doi.org/10.1016/j.jval.2014.04.007

91. Bessen T, Chen G, Street J, Eliott J, Karnon J, Keefe D, et al. What sort of follow-up services would Australian breast cancer survivors prefer if we could no longer offer long-term specialist-based care? A discrete choice experiment. Br J Cancer [Internet]. Nature Publishing Group; 2014;110:859–67. Available from: http://dx.doi.org/10.1038/bjc.2013.800

92. Bewtra M, Kilambi V, Fairchild AO, Siegel CA, Lewis JD, Johnson FR. Patient Preferences for Surgical Versus Medical Therapy for Ulcerative Colitis. Inflamm Bowel Dis [Internet]. 2014 [cited 2018 Jun 19];20:103–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24280881

93. Fraenkel L, Suter L, Cunningham CE, Hawker G. Understanding Preferences for Disease-Modifying Drugs in Osteoarthritis. Arthritis Care Res (Hoboken) [Internet]. 2014 [cited 2018 Jun 19];66:1186–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24470354

94. Goossens LMA, Utens CMA, Smeenk FWJM, Donkers B, Van Schayck OCP, Rutten-Van Mölken MPMH. Should i stay or should i go home? A latent class analysis of a discrete choice experiment on hospital-At-home. Value Heal [Internet]. Elsevier; 2014;17:588–96. Available from: http://dx.doi.org/10.1016/j.jval.2014.05.004

95. Hauber AB, Arellano J, Qian Y, González JM, Posner JD, Mohamed AF, et al. Patient preferences for treatments to delay bone metastases. Prostate. 2014;74:1488–97.

96. Havrilesky LJ, Secord AA, Ehrisman JA, Berchuck A, Valea FA, Lee PS, et al. Patient preferences in advanced or recurrent ovarian cancer. Cancer. 2014;120:3651–9.

97. Howard K, Gerard K, Adelson P, Bryce R, Wilkinson C, Turnbull D. Women's preferences for inpatient and outpatient priming for labour induction: A discrete choice experiment. BMC Health Serv Res. 2014;14:1–11.

98. Johnson P, Bancroft T, Barron R, Legg J, Li X, Watson H, et al. Discrete choice experiment to estimate breast cancer patients' preferences and willingness to pay for prophylactic granulocyte colony-stimulating factors. Value Heal [Internet]. Elsevier; 2014;17:380–9. Available from: http://dx.doi.org/10.1016/j.jval.2014.01.002

99. Kawata AK, Kleinman L, Harding G, Ramachandran S. Evaluation of Patient Preference and Willingness to Pay for Attributes of Maintenance Medication for Chronic Obstructive Pulmonary Disease (COPD). Patient - Patient-Centered Outcomes Res [Internet]. 2014;7:413–26. Available from: http://link.springer.com/10.1007/s40271-014-0064-1

100. Kløjgaard ME, Manniche C, Pedersen LB, Bech M, Søgaard R. Patient preferences for treatment of low back pain - A discrete choice experiment. Value Heal [Internet]. Elsevier; 2014;17:390–6. Available from: http://dx.doi.org/10.1016/j.jval.2014.01.005

101. Nafees B, Setyawan J, Lloyd A, Ali S, Hearn S, Sasane R, et al. Parent preferences regarding stimulant therapies for ADHD: a comparison across six European countries. Eur Child Adolesc Psychiatry. 2014;23:1189–200.

102. Ostermann J, Njau B, Brown DS, Mühlbacher A, Thielman N. Heterogeneous HIV testing preferences in an urban setting in tanzania: Results from a discrete choice experiment. PLoS One. 2014;9.

103. Poulos C, Hauber AB, González JM, Turpcu A. Patients' willingness to trade off between the duration and frequency of rheumatoid arthritis treatments. Arthritis Care Res. 2014;66:1008–15.

104. Smith M Lou, White CB, Railey E, Sledge GW. Examining and predicting drug preferences of patients with metastatic breast cancer: Using conjoint analysis to examine attributes of paclitaxel and capecitabine. Breast Cancer Res Treat. 2014;145:83–9.

105. Torbica A, Fattore G, Ayala F. Eliciting preferences to inform patient-centred policies: The case of psoriasis. Pharmacoeconomics. 2014;32:209–23.

106. Van Den Wijngaard L, Van Wely M, Dancet EAF, Van Mello NM, Koks CAM, Van Der Veen F, et al. Patients' preferences for gonadotrophin-releasing hormone analogs in in vitro fertilization. Gynecol Obstet Invest. 2014;78:16–21.

107. Van Haaren-ten Haken T, Pavlova M, Hendrix M, Nieuwenhuijze M, de Vries R, Nijhuis J. Eliciting preferences for key attributes of intrapartum care in the Netherlands. Birth. 2014;41:185–94.

108. Wijnen BFM, de Kinderen RJA, Colon AJ, Dirksen CD, Essers BAB, Hiligsmann M, et al. Eliciting patients' preferences for epilepsy diagnostics: A discrete choice experiment. Epilepsy Behav [Internet]. Elsevier Inc.; 2014;31:102–9. Available from: http://dx.doi.org/10.1016/j.yebeh.2013.11.029

109. Gerard K, Tinelli M, Latter S, Smith A, Blenkinsopp A. Patients' valuation of the prescribing nurse in primary care: A discrete choice experiment. Heal Expect. 2015;18:2223–35.

110. Krucien N, Le M, Ms V, Pelletier-fleury N. What are the patients ' preferences for the Chronic Care Model ? An application to the obstructive sleep apnoea syndrome. 2014;2536–48.

111. Bewtra M, Fairchild AO, Gilroy E, Leiman DA, Kerner C, Johnson FR, et al. Inflammatory bowel disease patients' willingness to accept medication risk to avoid future disease relapse. Am J Gastroenterol [Internet]. Nature Publishing Group; 2015;110:1675–81. Available from: http://dx.doi.org/10.1038/ajg.2015.321

112. Vries ST De, Vries FM De, Dekker T, Haaijer-ruskamp FM, De Vries ST, De Vries FM, et al. The role of patients' age on their preferences for choosing additional blood pressure-lowering drugs: A discrete choice experiment in patients with diabetes. PLoS One [Internet]. 2015;10:1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0139755

113. Bottger B, Thate-Waschke I-M, Bauersachs R, Kohlmann T, Wilke T. Preferences for anticoagulation therapy in atrial fibrillation: the patients' view. J Thromb Thrombolysis [Internet]. 2015;40:406–15. Available from: http://link.springer.com/10.1007/s11239-015-1263-x

114. De Bekker-Grob EW, Niers EJ, Van Lanschot JJB, Steyerberg EW, Wijnhoven BPL, Eva EWDB, et al. Patients' Preferences for Surgical Management of Esophageal Cancer: A Discrete Choice Experiment. World J Surg. Springer International Publishing; 2015;39:2492–9.

115. Howard K, Mph HI, Mhealthec M, Bbus GPS, Mph MPP. Men ' s preferences and trade-offs for prostate cancer screening : a discrete choice experiment. 2014;3123–35.

116. Beulen L, Grutters JPC, Faas BHW, Feenstra I, Groenewoud H, van Vugt JMG, et al. Women's and healthcare professionals' preferences for prenatal testing: A discrete choice experiment. Prenat Diagn. 2015;35:549–57.

117. Böttger B, Thate-Waschke I-M, Bauersachs R, Kohlmann T, Wilke T. Preferences for anticoagulation therapy in atrial fibrillation. J Thromb Thrombolysis. 2015;40:406–15.

118. Brown P, Panattoni L, Cameron L, Knox S, Ashton T, Tenbensel T, et al. Hospital sector choice and support for public hospital care in New Zealand : Results from a labeled discrete choice survey. J Health Econ [Internet]. Elsevier B.V.; 2015;43:118–27. Available from: http://dx.doi.org/10.1016/j.jhealeco.2015.06.004

119. Chamot E, Mulambia C, Kapambwe S, Shrestha S, Parham GP, Macwan'gi M, et al. Preference for human papillomavirus-based cervical cancer screening: Results of a choice-based conjoint study in Zambia. J Low Genit Tract Dis. 2015;19:119–23.

120. Chen L-C, Cheng L-J, Zhang Y, He X, Knaggs RD. Acupuncture or Low Frequency Infrared Treatment for Low Back Pain in Chinese Patients: A Discrete Choice Experiment. PLoS One [Internet]. 2015;10:e0126912. Available from: http://dx.plos.org/10.1371/journal.pone.0126912

121. Coxon D, Frisher M, Jinks C, Jordan K, Paskins Z, Peat G. The relative importance of perceived doctor's attitude on the decision to consult for symptomatic osteoarthritis: A choice-based conjoint analysis study. BMJ Open. 2015;5:1–10.

122. Danyliv A, Pavlova M, Gryga I, Groot W. Preferences for physician services in Ukraine : a discrete choice experiment. 2015;346–65.

123. Harris P, Whitty JA, Kendall E, Ratcliffe J, Wilson A, Littlejohns P, et al. The Australian public ' s preferences for emergency care alternatives and the in fl uence of the presenting context : a discrete choice experiment. 2015;1–12.

124. Laba T, Howard K, Rose J, Peiris D, Redfern J, Usherwood T, et al. Patient Preferences for a Polypill for the Prevention of Cardiovascular Diseases. 2015.

125. Larson E, Vail D, Mbaruku GM, Kimweri A, Freedman LP, Kruk ME. Moving Toward Patient-Centered Care in Africa : A Discrete Choice Experiment of Preferences for Delivery Care among 3 , 003 Tanzanian Women. 2015;1–12.

126. Mcateer A, Yi D, Watson V, Norwood P, Ryan M, Hannaford PC, et al. Exploring preferences for symptom management in primary care : 2015;478–88.

127. Berhane A, Enquselassie F. Patients' preferences for attributes related to health care services at hospitals in Amhara Region, northern Ethiopia: a discrete choice experiment. Patient Prefer Adherence [Internet]. 2015;9:1293–301. Available from: http://www.proxy.its.virginia.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=110257669&site=ehost-live&scope=site

128. Michaels-igbokwe C, Lagarde M, Cairns J, Initiative I, Terris-prestholt F. Designing a package of sexual and reproductive health and HIV outreach services to meet the heterogeneous preferences of young people

in Malawi : results from a discrete choice experiment. 2015;15–7.

129.Ostermann J, Njau B, Mtuy T, Brown DS, Mühlbacher A, Thielman N. One size does not fit all: HIV testing preferences differ among high-risk groups in Northern Tanzania. AIDS Care - Psychol Socio-Medical Asp AIDS/HIV [Internet]. 2015 [cited 2018 Jun 19];27:595–603. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25616562

130.Salloum RG, Maziak W, Hammond D, Nakkash R, Islam F, Cheng X, et al. Eliciting preferences for waterpipe tobacco smoking using a discrete choice experiment : implications for product regulation. 2015;1–8.

131.Sawamura K, Sano H, Rn MN. Japanese Public Long-Term Care Insured : Preferences for Future Long-Term Care Facilities , Including Relocation , Waiting Times , and Individualized Care. JMDA [Internet]. Elsevier; 2015;16:350.e9-350.e20. Available from: http://dx.doi.org/10.1016/j.jamda.2015.01.082

132.Brinovec RP, Warren E, Wittgens K, Dickmann P. Decision-making criteria among European patients : exploring patient preferences for primary care services. 2018;25:3–9.

133.Filby A, Smith AB, Carr LM. CONSUMER PREFERENCES FOR SCANNING MODALITY TO DIAGNOSE FOCAL LIVER LESIONS. 2015;2:27–35.

134.Whitty JA, Kendall E, Sav A, et al. Preferences for the delivery of community pharmacy services to help manage chronic conditions. Res Soc Adm Pharm [Internet]. Elsevier Inc; 2015;11:197–215. Available from: http://dx.doi.org/10.1016/j.sapharm.2014.06.007

135.Zickafoose JS, DeCamp LR, Prosser LA. Parents’ preferences for enhanced access in the pediatric medical home: a discrete choice experiment. JAMA Pediatr [Internet]. NIH Public Access; 2015 [cited 2018 Jun 19];169:358–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25643000

136.Hauber AB, Nguyen H, Posner J, Kalsekar I, Ruggles J. A discrete-choice experiment to quantify patient preferences for frequency of glucagon-like peptide-1 receptor agonist injections in the treatment of type 2 diabetes Original article A discrete-choice experiment to quantify patient preferences for frequen. 2016;7995.

137.Spinks J, Janda M, Soyer HP, Whitty JA. Consumer preferences for teledermoscopy screening to detect melanoma early. 2016;22:39–46.

138.Ramirez, M., Wu, S., Beale, E. Designing a Text Messaging Intervention to Improve Physical Activity Behavior Among Low-Income Latino Patients With Diabetes. Los Angeles, 2014–2015. Prev Chronic Dis 2016;13:160035. DOI: http://dx.doi.org/10.5888/pcd13.160035

139.Murchie P, Norwood PF, Pietrucin-materek M, Porteous T, Hannaford PC, Ryan M. Determining cancer survivors ’ preferences to inform new models of follow-up care. Nature Publishing Group; 2016;115:1495–503. Available from: http://dx.doi.org/10.1038/bjc.2016.352

140.Tang C, Xu J, Zhang M. The choice and preference for public- private health care among urban residents in China : evidence from a discrete choice experiment. BMC Health Serv Res [Internet]. BMC Health Services Research; 2016;1–11. Available from: http://dx.doi.org/10.1186/s12913-016-1829-0

141.Dong D, Ozdemir S, Bee YM, Toh S, Bilger M, Finkelstein E. Measuring High-Risk Patients ’ Preferences for Pharmacogenetic Testing to Reduce Severe Adverse Drug Reaction : A Discrete Choice Experiment. Value Heal [Internet]. Elsevier; 2016;19:767–75. Available from: http://dx.doi.org/10.1016/j.jval.2016.03.1837

142. Abiiro GA, Torbica A, Kwalamasa K. What factors drive heterogeneity of preferences for micro-health insurance in rural Malawi ? 2018;1172–83.

143. Minh H Van, Chung LH, Giang KB, Duc DM, Hinh ND, Mai VQ, et al. Potential Impact of Graphic Health Warnings on Cigarette Packages in Reducing Cigarette Demand and Smoking-Related Deaths in Vietnam. 2016;17:85–90.

144. Porteous T, Ryan M, Bond C, Watson M, Watson V. Managing Minor Ailments ; The Public ' s Preferences for Attributes of Community Pharmacies . A Discrete Choice Experiment. 2016;1–15.

145. Veldwijk J, Essers BAB, Lambooij MS, Dirksen CD, Smit HA, Wit GA De. Survival or Mortality : Does Risk Attribute Framing In fl uence Decision-Making Behavior in a Discrete Choice Experiment ? Value Heal [Internet]. Elsevier; 2016;19:202–9. Available from: http://dx.doi.org/10.1016/j.jval.2015.11.004

146. Wong SF, Sci BM, Norman R, Dunning TL, Ashley DM, Collins I, et al. A Discrete Choice Experiment to Examine the Preferences of Patients With Cancer and Their Willingness to Pay for Different Types of Health Care Appointments. 2016;14:311–9.

147. Ozawa S, Grewal S, Bridges JFP. Household Size and the Decision to Purchase Health Insurance in Cambodia : Results of a Discrete-Choice Experiment with Scale Adjustment. Appl Health Econ Health Policy. Springer International Publishing; 2016;14:195–204.

148. Czoli CD, Goniewicz M, Islam T, Kotnowski K, Hammond D. Consumer preferences for electronic cigarettes : results from a discrete choice experiment. 2016;30–7.

149. Veldwijk J, Lambooij MS, Kallenberg FGJ, Kranen HJ Van, Bredenoord AL, Dekker E, et al. Preferences for genetic testing for colorectal cancer within a population-based screening program : a discrete choice experiment. 2016;1:361–6.

150. Kotnowski K, Fong GT, Gallopel-morvan K, Islam T, Hammond D. Original investigation The Impact of Cigarette Packaging Design Among Young Females in Canada : Findings From a Discrete Choice Experiment. 2018;1348–56.

151. Kan HJ, de Bekker-Grob EW, Van Marion ES, Van Oijen GW, van Nieuwenhoven CA, Zhou C et al. Patients' preferences for treatment for dupuytren's disease: A discrete choice experiment. Plastic and Reconstructive Surgery. 2016;137(1), 165–173. doi:10.1097/PRS.0000000000001878

152. Marshall HS, Chen G, Clarke M, Ratcliffe J. Adolescent , parent and societal preferences and willingness to pay for meningococcal B vaccine : A Discrete Choice Experiment. Vaccine [Internet]. Elsevier Ltd; 2016;34:671–7. Available from: http://dx.doi.org/10.1016/j.vaccine.2015.11.075

153. Wang B, Chen G, Ratcliffe J, Haji H, Afzali A, Giles L, et al. Adolescent values for immunisation programs in Australia : A discrete choice experiment. 2017;28:1–14. Available from: http://dx.doi.org/10.1371/journal.pone.0181073

154. Determann D, Lambooij MS, Steyerberg EW, Bekker-grob EW De, Wit GA De. Impact of Survey Administration Mode on the Results of a Health-Related Discrete Choice Experiment : Online and Paper Comparison. Value Heal [Internet]. Elsevier Inc.; 2017;20:953–60. Available from: http://dx.doi.org/10.1016/j.jval.2017.02.007

155. Wright SJ, Ulph F, Dharni N. Eliciting Preferences for Information Provision in Newborn Bloodspot Screening Programs. Value Health [Internet]. Elsevier Inc.; 2017;20:651–61. Available from: http://dx.doi.org/10.1016/j.jval.2016.11.007

156. Carlin CS, Higuera L, Anderson S. Improving Patient-Centered Care by Assessing Patient Preferences for Multiple Sclerosis Disease-Modifying Agents : A Stated-Choice Experiment. 2017;1–7.

157. Eliasson L, Freitas HM De, Dearden L, Calimlim B, Lloyd AJ. Patients ' Preferences for the Treatment of Metastatic Castrate-resistant Prostate Cancer : A Discrete Choice Experiment. Clin Ther [Internet]. Elsevier HS Journals, Inc.; 2017;39:723–37. Available from: http://dx.doi.org/10.1016/j.clinthera.2017.02.009

158. Marshall DA, Gonzalez JM, Macdonald K V, Johnson FR. Estimating Preferences for Complex Health Technologies : Lessons Learned and Implications for Personalized Medicine. Value Heal [Internet]. Elsevier Inc.; 2017;20:32–9. Available from: http://dx.doi.org/10.1016/j.jval.2016.08.737

159. Cranen K, Groothuis-Oudshoorn CG, Vollenbroek-Hutten MM, IJzerman MJ. Toward Patient-Centered Telerehabilitation Design: Understanding Chronic Pain Patients' Preferences for Web-Based Exercise Telerehabilitation Using a Discrete Choice Experiment. J Med Internet Res [Internet]. JMIR Publications Inc.; 2017 [cited 2018 Jun 19];19:e26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28108429

160. Determann D, Lambooij MS, Gyrd-hansen D, Bekker-grob EW De, Steyerberg EW, Heldoorn M, et al. Personal health records in the Netherlands : potential user preferences quantified by a discrete choice experiment. 2017;24:529–36.

161. Flood EM, Bell KF, Cruz MC De, Ginchereau- FM. Patient preferences for diabetes treatment attributes and drug classes. 2017;7995.

162. Deen WK Van, Nguyen D, Duran NE. Value redefined for inflammatory bowel disease patients : a choice-based conjoint analysis of patients ' preferences. Springer International Publishing; 2017;455–65.

163. Miners AH, Llewellyn CD, Cooper VL, Youssef E, Pollard AJ, Lagarde M, et al. A discrete choice experiment to assess people living with HIV ' s (PLWHIV ' s) preferences for GP or HIV clinic appointments. 2016;1–7.

164. Hauber AB, Mohamed AF, Gonzalez JM, Fairchild O, Zelt SC, Graff O, et al. Benefit-risk tradeoff preferences for chronic hand eczema treatments. 2017;6634.

165. Mühlbacher AC, Bridges JFP, Bethge S, Dintsios C-M, Schwalm A, Gerber-Grote A, et al. Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. Eur J Heal Econ [Internet]. 2017;18:155–65. Available from: http://link.springer.com/10.1007/s10198-016-0763-8

166. Marshall T, Pugh A, Fairchild A, Hass S. Patient Preferences for Device-Aided Treatments Indicated for Advanced Parkinson Disease. Value Heal [Internet]. Elsevier Inc.; 2017;20:1383–93. Available from: http://dx.doi.org/10.1016/j.jval.2017.06.001

167. Seanehia J, Treibich C, Holmberg C, Müller-Nordhorn J, Casin V, Raude J, et al. Quantifying population preferences around vaccination against severe but rare diseases: A conjoint analysis among French university students, 2016. Vaccine [Internet]. 2017 [cited 2018 Jul 12];35:2676–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28408120

168. Zipursky RB, Cunningham CE, Stewart B, Rimas H, Cole E, Vaz SM. Characterizing outcome preferences in patients with psychotic disorders: a discrete choice conjoint experiment. Schizophr Res [Internet]. 2017 [cited 2018 Jul 12];185:107–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28038919

169. Alayli-Goebbels AFG, Dellaert BGC, Knox SA, Ament AJHA, Lakerveld J, Bot SDM, et al. Consumer preferences for health and nonhealth outcomes of health promotion: Results from a discrete choice experiment. Value Heal. 2013;16:114–23.

170. Bailey H. Results from a preliminary study to develop the quality adjustments for quality adjusted life year values for Trinidad and Tobago. West Indian Med J. 2013;62:543–7.

171. Gu NY, Botteman MF, Gerber RA, Ji X, Postema R, Wan Y, et al. Eliciting health state utilities for Dupuytren's contracture using a discrete choice experiment. Acta Orthop. 2013;84:571–8.

172. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl Health Econ Health Policy. 2013;11:287–98.

173. Prosser LA, Payne K, Rusinak D, Shi P, Messonnier M. Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages. Pharmacoeconomics. 2013;31:305–15.

174. Howard K, Salkeld GP, Patel MI, Mann GJ, Pignone MP. Men's preferences and trade-offs for prostate cancer screening: A discrete choice experiment. Heal Expect. 2015;18:3123–35.

175. Notten KJB, Essers BA, Weemhoff M, Rutten AGH, Donners J, Gestel I Van. Do patients prefer mesh or anterior colporrhaphy for primary correction of anterior vaginal wall prolapse : a labelled discrete choice experiment. 2015;873–80.

176. Wijngaard L Van Den, Rodijk ICM, Veen F Van Der, Erven MHWG, Koks CAM, Verhoeve HR, et al. Patient preference for a long-acting recombinant FSH product in ovarian hyperstimulation in IVF : a discrete choice experiment. 2015;30:331–7.

177. Kauf TL, Yang J, Kimball AB, Sundaram M, Okun M, Mulani P, et al. Psoriasis patients ' willingness to accept side-effect risks for improved treatment efficacy treatment efficacy. 2015;6634.

178. Miller FA, Hayeems RZ, Bombard Y, Cressman C, Barg CJ. Public Perceptions of the Bene fi ts and Risks of Newborn Screening. 2018;136.

179. Mohamed AF, González JM, Fairchild A. Patient Benefit-Risk Tradeoffs for Radioactive Iodine-Refractory Differentiated Thyroid Cancer Treatments. 2015;2015.

180. Feliciano R, Ponte C, Botella M, Lizán L. Patients ' and physicians ' preferences for type 2 diabetes mellitus treatments in Spain and Portugal : a discrete choice experiment. 2015;1443–58.

181. Christin J, Mu AC. Chronic pain patients ' treatment preferences : a discrete-choice experiment. 2015;613–28.

182. Ngorsuraches S, Nawanukool K, Petcharamanee K, Poopantrakool U. Parents ' preferences and willingness-to-pay for human papilloma virus vaccines in Thailand. J Pharm Policy Pract [Internet]. Journal of Pharmaceutical Policy and Practice; 2015; Available from: http://dx.doi.org/10.1186/s40545-015-0040-8

183. Ngorsuraches S, Thongkeaw K. Patients ' preferences and willingness - to - pay for postmenopausal hormone receptor - positive , HER2 - negative advanced breast cancer treatments after failure of standard treatments. Springerplus. Springer International Publishing; 2015;

184. Jarvik GP, Hoch J, Pharmd DV. CMAJ Societal preferences for the return of incidental findings from clinical genomic sequencing: a discrete-choice experiment. 2015;187.

185. Ryan M. Gaining pounds by losing pounds : preferences for lifestyle interventions to reduce obesity. 2015;161–82.

186. Finkelstein E, Malhotra C, Chay J, Ozdemir S, Chopra A, Kanesvaran R. Impact of Treatment Subsidies and Cash Payouts on Treatment Choices at the End of Life. Value Heal [Internet]. Elsevier; 2016;19:788–94.

Available from: http://dx.doi.org/10.1016/j.jval.2016.02.015

187. Shiroiwa T, Ikeda S, Noto S, Igarashi A. Comparison of Value Set Based on DCE and / or TTO Data : Scoring for EQ-5D-5L Health States in Japan. Value Heal [Internet]. Elsevier; 2016;19:648–54. Available from: http://dx.doi.org/10.1016/j.jval.2016.03.1834

188. Versteegh MM, Vermeulen KM, Evers SMAA, Wit GA De, Prenger R, Stolk EA. Dutch Tariff for the Five-Level Version of EQ-5D. Value Heal [Internet]. Elsevier; 2016;19:343–52. Available from: http://dx.doi.org/10.1016/j.jval.2016.01.003

189. Mulhern B, Shah K, Janssen MFB, Longworth L, Ibbotson R. Valuing Health Using Time Trade-Off and Discrete Choice Experiment Methods : Does Dimension Order Impact on Health State Values ? Value Heal [Internet]. Elsevier; 2016;19:210–7. Available from: http://dx.doi.org/10.1016/j.jval.2015.11.005

190. Wanishayakorn T, Sornlertlumvanich K, Ngorsuraches S. Bene fi t – risk assessment of HMG-CoA reductase inhibitors (statins): a discrete choice experiment. 2016;

191. Cross J, Yang J, Johnson FR, Quiroz J, Dunn J, Raspa M, et al. Fragile X Syndrome. 2016;37:71–9.

192. De Bekker-Grob EW, Rose JM, Donkers B, Essink-Bot ML, Bangma CH, Steyerberg EW. Men’s preferences for prostate cancer screening: A discrete choice experiment. Br J Cancer [Internet]. Nature Publishing Group; 2013;108:533–41. Available from: http://dx.doi.org/10.1038/bjc.2013.5

193. Dwight Johnson M, Apesoa-Varano C, Hay J, Unutzer J, Hinton L. Depression treatment preferences of older white and Mexican origin men. Gen Hosp Psychiatry [Internet]. Elsevier B.V.; 2013;35:59–65. Available from: http://dx.doi.org/10.1016/j.genhosppsych.2012.08.003

194. Kjær T, Bech M, Kronborg C, Mørkbak MR. Public preferences for establishing nephrology facilities in Greenland: Estimating willingness-to-pay using a discrete choice experiment. Eur J Heal Econ. 2013;14:739–48.

195. Knox SA, Viney RC, Gu Y, Hole AR, Fiebig DG, Street DJ, et al. The effect of adverse information and positive promotion on women’s preferences for prescribed contraceptive products. Soc Sci Med [Internet]. Elsevier Ltd; 2013;83:70–80. Available from: http://dx.doi.org/10.1016/j.socscimed.2012.12.025

196. Lynn FA, Crealey GE, Alderdice FA, Mcelnay JC. Preferences for a third-trimester ultrasound scan in a low-risk obstetric population : a discrete choice experiment. 2013;892–903.

197. Agyei-Baffour P, Boahemaa MY, Addy EA. Contraceptive preferences and use among auto artisanal workers in the informal sector of Kumasi, Ghana: A discrete choice experiment. Reprod Health [Internet]. 2015;12:1–11.

198. Arbiol J, Yabe M, Nomura H, Borja M, Gloriani N, Yoshida SI. Using discrete choice modeling to evaluate the preferences and willingness to pay for leptospirosis vaccine. Hum Vaccines Immunother. 2015;11:1046–56.

199. Poon J, Curtis SE, Boye KS. Evaluating preferences for profiles of GLP-1 receptor agonists among injection-naïve type 2 diabetes patients in the UK. 2015;1611–22.

200. Groenewoud S, Van Exel NJA, Bobinac A, Berg M, Huijsman R, Stolk EA. What Influences Patients’ Decisions When Choosing a Health Care Provider? Measuring Preferences of Patients with Knee Arthrosis, Chronic Depression, or Alzheimer’s Disease, Using Discrete Choice Experiments. Health Serv Res [Internet]. 2015 [cited 2018 Jun 19];50:1941–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26768957

201. Hauber AB, Tunceli K, Gantz I, Brodovicz KG, Alexander CM, Davies MJ, et al. ORIGINAL RESEARCH A Survey of Patient Preferences for Oral Antihyperglycemic Therapy in Patients with Type 2 Diabetes Mellitus. 2015;75–84.

202. Hechmati G, Hauber AB, Arellano J, Mohamed AF, Qian Y. Patients ' preferences for bone metastases treatments in France , Germany and the United Kingdom. 2015;21–8.

203. Kistler CE, Hess TM, Howard K, Pignone MP, Crutchfield TM, Hawley ST, et al. Older adults ' preferences for colorectal cancer-screening test attributes and test choice. 2015;1005–16.

204. Kromer C, Schaarschmidt M, Schmieder A, Herr R. Patient Preferences for Treatment of Psoriasis with Biologicals : A Discrete Choice Experiment. 2015;1–15.

205. Krucien N, Gafni A, Pelletier-fleury N. EMPIRICAL TESTING OF THE EXTERNAL VALIDITY OF A DISCRETE CHOICE EXPERIMENT TO DETERMINE PREFERRED TREATMENT OPTION : THE CASE OF SLEEP APNEA. 2015;965:951–65.

206. Lambooij MS, Harmsen IA, Veldwijk J, Melker H De, Mollema L, Weert YWM Van, et al. Consistency between stated and revealed preferences : a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. 2015;1–8.

207. Levitan B, Markowitz M, Mohamed AF, Johnson FR, Alphs L, Citrome L, et al. Patients' Preferences Related to Benefits, Risks, and Formulations of Schizophrenia Treatment. Psychiatr Serv [Internet]. 2015 [cited 2018 Jun 19];66:719–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25772762

208. Malhotra C, Farooqui MA, Kanesvaran R, Bilger M, Finkelstein E. Comparison of preferences for end-of-life care among patients with advanced cancer and their caregivers : A discrete choice experiment. 2015;

209. Johnson FR, Farah A, Reed MF, Balp M. Preferences and Stated Adherence for Antibiotic Treatment of Cystic Fibrosis Pseudomonas Infections. Preferences and Stated Adherence for Antibiotic Treatment of Cystic Fibrosis Pseudomonas Infections. 2015.

210. Mühlbacher AC, Bethge S. Reduce Mortality Risk Above All Else: A Discrete-Choice Experiment in Acute Coronary Syndrome Patients. Pharmacoeconomics. 2014;33:71–81.

211. Mühlbacher AC, Bethge S. Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. Eur J Heal Econ [Internet]. 2015;16:657–70. Available from: http://link.springer.com/10.1007/s10198-014-0622-4

212. Nafees B, Lloyd A, Elkin E, Porret T, Nafees B, Lloyd A, et al. To explore preferences and willingness to pay for attributes regarding stoma appliances amongst patients in the UK , France and Germany Original article To explore preferences and willingness to pay for attributes regarding stoma appliances amongst patien. 2015;7995.

213. Pacou M, Basso F, Gore C, Hass B, Taieb V, Cognet M, et al. Patient and physician preferences for the treatment of chronic hepatitis C virus infections : does the perspective matter ? 2015;1063–8.

214. Pinder B, Lloyd AJ, Marley J. Patient preferences and willingness to pay for innovations in intermittent self-catheters. 2015;381–8.

215. Pisa G, Eichmann F, Hupfer S. Assessing patient preferences in heart failure using conjoint methodology. Patient Preference and Adherence, 2015;1233–41.

216. Salampessy BH, Veldwijk J, Schuit AJ, Wit GA De, Lambooij MS. The Predictive Value of Discrete Choice Experiments in Public Health : An Exploratory Application. Patient - Patient-Centered Outcomes Res [Internet]. Springer International Publishing; 2015;521–9. Available from:

http://dx.doi.org/10.1007/s40271-015-0115-2

217. Salloum RG, Rn CWA, Kohler RE, Ba AKK, Oh L, Rn KAW. Original investigation Assessing Preferences for a University-Based Smoking Cessation Program in Lebanon : A Discrete Choice Experiment. 2015;580–5.

218. Schaarschmidt M, Kromer C, Herr R, Schmieder A. Patient Preferences for Biologicals in Psoriasis : Top Priority of Safety for Cardiovascular Patients. 2015;1–14.

219. Veldwijk J, Lambooij MS, Til JA Van, Smit A, Td GA, Td IF De, et al. Patient Education and Counseling Words or graphics to present a Discrete Choice Experiment : Does it matter ? 2015;98:1376–84.

220. Uemura H, Matsubara N, Kimura G, Yamaguchi A, Ledesma DA, Dibonaventura M, et al. Patient preferences for treatment of castration-resistant prostate cancer in Japan : a discrete-choice experiment. BMC Urol [Internet]. BMC Urology; 2016;2–4. Available from: http://dx.doi.org/10.1186/s12894-016-0182-2

221. Mühlbacher A, Bethge S. What matters in type 2 diabetes mellitus oral treatment? A discrete choice experiment to evaluate patient preferences. Eur J Heal Econ [Internet]. 2016;17:1125–40. Available from: http://link.springer.com/10.1007/s10198-015-0750-5

222. Lee J, Kim K, Shin Y, Young H, Ji E, Kim S, et al. Gynecologic Oncology Treatment preferences of advanced ovarian cancer patients for adding bevacizumab to fi rst-line therapy. Gynecol Oncol [Internet]. Elsevier B.V.; 2016;143:622–7. Available from: http://dx.doi.org/10.1016/j.ygyno.2016.10.021

223. Brown ZS, Kramer RA, Ocan D, Oryema C. Household perceptions and subjective valuations of indoor residual spraying programmes to control malaria in northern Uganda. Infect Dis Poverty [Internet]. Infectious Diseases of Poverty; 2016;1–13. Available from: http://dx.doi.org/10.1186/s40249-016-0190-1

224. Mühlbacher A, Bethge S. First and Foremost Battle the Virus : Eliciting Patient Preferences in Antiviral Therapy for Hepatitis C Using a Discrete Choice Experiment. Value Heal [Internet]. Elsevier; 2016;19:776–87. Available from: http://dx.doi.org/10.1016/j.jval.2016.04.007

225. Franco MR. Smallest worthwhile effect of exercise programs to prevent falls among older people : estimates from benefit – harm trade-off and discrete choice methods. 2018;806–12.

226. Hazlewood GS, Bombardier C, Tomlinson G, Thorne C, Bykerk VP, Thompson A, et al. Original article Treatment preferences of patients with early rheumatoid arthritis : a discrete-choice experiment. 2018;1959–68.

227. Shafrin J, Bruno A, Macewan JP, Campinha-bacote A, Trocio J, Shah M, et al. Physician and Patient Preferences for Nonvalvular Atrial Fibrillation Therapies. Value Heal [Internet]. Elsevier; 2016;19:451–9. Available from: http://dx.doi.org/10.1016/j.jval.2016.01.001

228. Dijk JD Van, Groothuis-oudshoorn CGM, Marshall DA, Ijzerman MJ. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders ' Risk Tolerance for Hip Replacement Surgery. Value Heal [Internet]. Elsevier; 2016;19:316–22. Available from: http://dx.doi.org/10.1016/j.jval.2015.12.020

229. Ride J, Lancsar E. Women ' s Preferences for Treatment of Perinatal Depression and Anxiety : A Discrete Choice Experiment. 2016;1–23.

230. Morel T, Aymé S, Cassiman D, Simoens S, Morgan M, Vandebroek M. Quantifying benefit-risk preferences for new medicines in rare disease patients and caregivers. Orphanet J Rare Dis [Internet]. Orphanet Journal of Rare Diseases; 2016;1–12. Available from: http://dx.doi.org/10.1186/s13023-016-0444-9

231. Holmes EAF, Morrison VL, Hughes DA. What in fluences persistence with medicines ? A multinational discrete choice experiment of 2549 patients. PharmacoEconomics,2016;522–31.

232. Craig BM, Mitchell SA, Ct ABSTRA. Examining the Value of Menopausal Symptom Relief Among US Women. Value Heal [Internet]. Elsevier; 2016;19:158–66. Available from: http://dx.doi.org/10.1016/j.jval.2015.11.002

233. Niroomand N, Jenkins GP. Estimating the Value of Life , Injury , and Travel Time Saved Using a Stated Preference Framework. Accid Anal Prev [Internet]. Elsevier Ltd; 2016;91:216–25. Available from: http://dx.doi.org/10.1016/j.aap.2016.03.004

234. O'Hara NN, Slobogean GP, Mohammadi T. Are patients willing to pay for total shoulder arthroplasty? Evidence from a discrete choice experiment. Canadian Journal of Surgery. 2016;59(2):107-112. doi:10.1503/cjs.011915.

235. Qian Y, Arellano J, Hauber AB, Mohamed AF, Marcos J, Guy G, et al. Patient , Caregiver , and Nurse Preferences for Treatments for Bone Metastases from Solid Tumors. Patient - Patient-Centered Outcomes Res. Springer International Publishing; 2016;9:323–33.

236. Kerr C, Lloyd EJ, Kosmas CE, Smith HT, Cooper JA, Johnston K, et al. Health-related quality of life in Parkinson ' s : impact of ' off ' time and stated treatment preferences. Qual Life Res. Springer International Publishing; 2016;25:1505–15.

237. Poulos C, Kinter E, John JY, Posner J, Reder AT. Patient Preferences for Injectable Treatments for Multiple Sclerosis in the United States : A Discrete-Choice Experiment. Patient - Patient-Centered Outcomes Res. Springer International Publishing; 2016;9:171–80.

238. Arx L Von, Johnson FR, Mørkbak MR, Kjær T. Be Careful What You Ask For : Effects of Bene fit Descriptions on Diabetes Patients ' Benefit-Risk Tradeoff Preferences. 2017;20:670–8.

239. Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. AN AUSTRALIAN DISCRETE CHOICE EXPERIMENT TO VALUE EQ-5D HEALTH STATES. Health Econ [Internet]. 2014 [cited 2018 Jun 19];23:729–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23765787

240. Eckert J, Lichters M, Piedmont S, Vogt B, Robra BP. Case vignettes based on EQ-5D to elicit stated preferences for health services utilization from the insurees' perspective. BMC Health Serv Res [Internet]. BMC Health Services Research; 2015;15:1–11. Available from: http://dx.doi.org/10.1186/s12913-015-1143-2

241. Craig BM, Brown DS, Reeve BB. The Value Adults Place on Child Health and Functional Status. Value Health [Internet]. NIH Public Access; 2015 [cited 2018 Jun 19];18:449–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26091599

242. Haagsma JA, Noordhout CM De, Polinder S, Vos T, Havelaar AH, Cassini A, et al. Assessing disability weights based on the responses of 30 , 660 people from four European countries. 2015;1–15.

243. Robinson A, Spencer A, Moffatt P. A Framework for Estimating Health State Utility Values within a Discrete Choice Experiment : Modeling Risky Choices. 2015;341–50.

244. Rosato R, Testa S, Oggero A, Molinengo G, Bertolotto A. Quality of life and patient preferences : identification of subgroups of multiple sclerosis patients. Qual Life Res [Internet]. Springer International Publishing; 2015;2173–82. Available from: http://dx.doi.org/10.1007/s11136-015-0952-4

245. Rowen D, Brazier J, Hout B Van. A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. 2015;328–40.

246. Scalone L, Stalmeier PFM, Milani S, Krabbe PFM. Values for health states with different life durations. Eur J Heal Econ [Internet]. Springer Berlin Heidelberg; 2015;917–25. Available from: http://dx.doi.org/10.1007/s10198-014-0634-0

247. Craig BM, Brown DS, Reeve BB. Valuation of Child Behavioral Problems from the Perspective of US Adults. Med Decis Making. 2017;36:199–209.

248. Norman R, Kemmler G, Viney R, Pickard AS, Gamper E, Holzner B, et al. Order of Presentation of Dimensions Does Not Systematically Bias Utility Weights from a Discrete Choice Experiment. Value Heal [Internet]. Elsevier; 2016;19:1033–8. Available from: http://dx.doi.org/10.1016/j.jval.2016.07.003

249. Ock M, Lee JY, Oh I, Park H, Yoon S, Jo M. Disability Weights Measurement for 228 Causes of Disease in the Korean Burden of Disease Study 2012. 2016;3:129–38.

250. Rowen D, Brazier J, Mukuria C, Keetharuth A, Hole AR, Tsuchiya A, et al. Eliciting Societal Preferences for Weighting QALYs for Burden of Illness and End of Life. 2016;210–22.

251. Mulhern B, Norman R, Lorgelly P, Lancsar E, Ratcliffe J, Brazier J, et al. Is Dimension Order Important when Valuing Health States Using Discrete Choice Experiments Including Duration ? Pharmacoeconomics. Springer International Publishing; 2017;35:439–51.

252. Ageyi-Baffour P, Rominski S, Nakua E, Gyakobo M, Lori JR. Factors that influence midwifery students in Ghana when deciding where to practice: A discrete choice experiment. BMC Med Educ. 2013;13:2–7.

253. Keuffel E, Jaskiewicz W, Paphassarang C, Tulenko K. Net costs of health worker rural incentive packages: An example from the lao people's democratic republic. Med Care. 2013;51:985–91.

254. Lagarde M, Pagaiya N, Tangcharoensathian V, Blaauw D. ONE SIZE DOES NOT FIT ALL: INVESTIGATING DOCTORS' STATED PREFERENCE HETEROGENEITY FOR JOB INCENTIVES TO INFORM POLICY IN THAILAND. Health Econ [Internet]. 2013 [cited 2018 Jun 19];22:1452–69. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23349119

255. Rao KD, Ryan M, Shroff Z, Vujicic M, Ramani S, Berman P. Rural clinician scarcity and job preferences of doctors and nurses in India: A Discrete Choice Experiment. PLoS One. 2013;8.

256. Scott A, Witt J, Humphreys J, Joyce C, Kalb G, Jeon SH, et al. Getting doctors into the bush: General practitioners' preferences for rural location. Soc Sci Med [Internet]. Elsevier Ltd; 2013;96:33–44. Available from: http://dx.doi.org/10.1016/j.socscimed.2013.07.002

257. Pedersen LB, Gyrd-Hansen D. Preference for practice: A Danish study on young doctors' choice of general practice using a discrete choice experiment. Eur J Heal Econ. 2014;15:611–21.

258. Holte J, Helgheim J, Kjaer T, Abelsen B, Abel J. Social Science & Medicine The impact of pecuniary and non-pecuniary incentives for attracting young doctors to rural general practice. Soc Sci Med [Internet]. Elsevier Ltd; 2015;128:1–9. Available from: http://dx.doi.org/10.1016/j.socscimed.2014.12.022

259. Honda A, Vio F. Incentives for non-physician health professionals to work in the rural and remote areas of Mozambique — a discrete choice experiment for eliciting job preferences. Hum Resour Health. 2015; Apr 26;13:23.

260. Robyn PJ, Shroff Z, Zang OR, Kingue S, Djienouassi S, Kouontchou C, et al. Original Article Addressing health workforce distribution concerns : a discrete choice experiment to develop rural retention strategies in Cameroon. 2015;4:169–80. Available from: http://dx.doi.org/10.15171/ijhpm.2015.27

261. Scott A, Witt J, Duffield C. What do nurses and midwives value about their jobs ? Results from a discrete choice experiment. 2015;20:31–8.

262. Song K, Scott A, Sivey P, Meng Q. Improving Chinese primary care providers ’ recruitment and retention : a discrete choice experiment. 2018;68–77.

263. Gallego G, Dew A, Lincoln M, Bundy A, Chedid RJ, Bulkeley K, et al. Should I stay or should I go ? Exploring the job preferences of allied health professionals working with people with disability in rural Australia. Human Resources for Health; 2015;1–13.

264. Smitz M, Witter S, Lemiere C, Hoang- P, Eozenou V, Lievens T, et al. Understanding Health Workers ’ Job Preferences to Improve Rural Retention in Timor-Leste : Findings from a Discrete Choice Experiment. 2016;1–14.

265. Shiratori S, Agyekum EO, Shibanuma A, Oduro A, Okawa S, Enuameh Y, et al. Motivation and incentive preferences of community health officers in Ghana : an economic behavioral experiment approach. Hum Resour Health [Internet]. Human Resources for Health; 2016; Available from: http://dx.doi.org/10.1186/s12960-016-0148-1

266. Efendi F, Chen C, Nursalam N, Wachyu N, Andriyani F, Kurniati A, et al. How to attract health students to remote areas in Indonesia : a discrete choice experiment. 2016;430–45.

267. Cleland J, Johnston P, Watson V, Krucien N, Ska D. trainees ’ desires What do UK doctors in training value in a post ? A discrete choice experiment. 2016;189–202.

268. Takemura T, Kielmann K, Blaauw D. Job preferences among clinical officers in public sector facilities in rural Kenya : a discrete choice experiment. Hum Resour Health [Internet]. Human Resources for Health; 2016;1–10. Available from: http://dx.doi.org/10.1186/s12960-015-0097-0

269. Linley WG, Hughes DA. Decision-makers’ preferences for approving new medicines in wales: A discrete-choice experiment with assessment of external validity. Pharmacoeconomics. 2013;31:345–55.

270. Norman R, Hall J, Street D, Viney R. EFFICIENCY AND EQUITY: A STATED PREFERENCE APPROACH. Health Econ [Internet]. Wiley-Blackwell; 2013 [cited 2018 Jun 19];22:568–81. Available from: http://doi.wiley.com/10.1002/hec.2827

271. Sadique MZ, Devlin N, Edmunds WJ, Parkin D. The Effect of Perceived Risks on the Demand for Vaccination: Results from a Discrete Choice Experiment. PLoS One. 2013;8.

272. Terris-Prestholt F, Hanson K, MacPhail C, Vickerman P, Rees H, Watts C. How much demand for new HIV prevention technologies can we really expect? Results from a discrete choice experiment in South Africa. PLoS One. 2013;8.

273. Skedgel C, Wailoo A, Akehurst R. Societal Preferences for Distributive Justice in the Allocation of Health Care Resources : A Latent Class Discrete Choice Experiment. 2015;94–105.

274. Mentzakis E, Paolucci F, Rubicko G. Priority setting in the Austrian healthcare system: results from a discrete choice experiment and implications for mental health. J Ment Health Policy Econ [Internet]. 2014 [cited 2018 Jun 19];17:61–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25153094

275. Michaels-Igbokwe C, Terris-Prestholt F, Lagarde M, Chipeta E, Cairns J, Mayhew S, et al. Young People’s Preferences for Family Planning Service Providers in Rural Malawi: A Discrete Choice Experiment. PLoS One [Internet]. 2015;10:1–18. Available from: http://dx.doi.org/10.1371/journal.pone.0143287

276. Bosworth R, Cameron TA, DeShazo JR. Willingness to pay for public health policies to treat illnesses. J Health Econ [Internet]. Elsevier B.V.; 2015;39:74–88. Available from:

http://dx.doi.org/10.1016/j.jhealeco.2014.10.004

277. Harrison M, Marra C, Shojania K, Bansback N. Original article Societal preferences for rheumatoid arthritis treatments : evidence from a discrete choice experiment. 2018;1816–25.

278. Honda A, Ryan M, Niekerk R Van, Mcintyre D. Improving the public health sector in South Africa : eliciting public preferences using a discrete choice experiment. 2018;600–11.

279. Luyten J, Kessels R, Goos P, Beutels P. Public Preferences for Prioritizing Preventive and Curative Health Care Interventions : A Discrete Choice Experiment. Value Heal [Internet]. Elsevier; 2015;18:224–33. Available from: http://dx.doi.org/10.1016/j.jval.2014.12.007

280. Paolucci F, Mentzakis E, Defechereux T, Niessen W., Equity and efficiency preferences of health policy makers in China—a stated preference analysis, Health Policy and Planning, Volume 30, Issue 8, 1 October 2015, Pages 1059–1066, https://doi.org/10.1093/heapol/czu123.

281. Severin F, Hess W, Schmidtke J, Mühlbacher A, Rogowski W. Value judgments for priority setting criteria in genetic testing : A discrete choice experiment. Health Policy (New York) [Internet]. Elsevier Ireland Ltd; 2015;119:164–73. Available from: http://dx.doi.org/10.1016/j.healthpol.2014.04.013

282. Shah KK, Tsuchiya A, Wailoo AJ. Social Science & Medicine Valuing health at the end of life : A stated preference discrete choice experiment. Soc Sci Med [Internet]. Elsevier Ltd; 2015;124:48–56. Available from: http://dx.doi.org/10.1016/j.socscimed.2014.11.022

283. Skedgel, C., Wailoo, A.J., Akehurst, R.L. Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. Medical decision-making : an international journal of the Society for Medical Decision-making, 2015, 35 1, 94-105.

284. Wetering L, Exel J, Bobinac A, Brouwer WBF. Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment. Pharmacoeconomics. Springer International Publishing; 2015;33:1289–300.

285. Whitty JA, Ratcliffe J, Kendall E, Burton P, Wilson A, Littlejohns P, et al. Prioritising patients for bariatric surgery : building public preferences from a discrete choice experiment into public policy. BMJ Open. 2015; 5:e008919. doi: 10.1136/bmjopen-2015-008919

286. Determann D, Korfage IJ, Fagerlin A, Steyerberg EW, Bliemer MC, Voeten HA, et al. Public preferences for vaccination programmes during pandemics caused by pathogens transmitted through respiratory droplets – a discrete choice experiment in four European countries , 2013. 2016;1–13.

287. Shiroiwa T, Saito S, Shimozuma K, Kodama S. Societal Preferences for Interventions with the Same Efficiency : Assessment and Application to Decision-making. Appl Health Econ Health Policy. Springer International Publishing; 2016;14:375–85.

288. Herman PM, Ingram M, Cunningham CE, Rimas H, Murrieta L, Schachter K, et al. ORIGINAL RESEARCH ARTICLE A Comparison of Methods for Capturing Patient Preferences for Delivery of Mental Health Services to Low-Income Hispanics Engaged in Primary Care. Patient - Patient-Centered Outcomes Res. Springer International Publishing; 2016;9:293–301.

289. Howard K, Jan S, Rose JM, Wong G, Craig JC, Irving M, et al. Preferences for Policy Options for Deceased Organ Donation for Transplantation : A Discrete Choice Experiment. 2016;100.

290. Becker MPE, Sc BA, Sc M, Christensen BK, Ph D, Cunningham CE, et al. Preferences for Early Intervention Mental Health Services : A Discrete-Choice Conjoint Experiment. 2016;

291.Ng V, Sargeant JM. Prioritizing Zoonotic Diseases : Differences in Perspectives Between Human and Animal Health Professionals in North America. PloS One. 2016;196–211.

292.Mühlbacher AC, Bethge S, Reed SD, Schulman KA. Patient Preferences for Features of Health Care Delivery Systems: A Discrete Choice Experiment. Health Serv Res [Internet]. 2016 [cited 2018 Jun 19];51:704–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26255998

293.Lock J, Urhan G, Peters M, Meijer K, Brons P. Facilitating the implementation of pharmacokinetic-guided dosing of prophylaxis in haemophilia care by discrete choice experiment. 2016;1–10.

294.Ozdemir S, Wong TT, Allingham RR, Finkelstein EA. Predicted patient demand for a new delivery system for glaucoma medicine. Medicine (Baltimore) [Internet]. 2017 [cited 2018 Jun 19];96:e6626. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28403115

295.Wetering EJ Van De, Exel J Van, Brouwer WBF. The Challenge of Conditional Reimbursement : Stopping Reimbursement Can Be More Dif fi cult Than Not Starting in the First Place ! Value Heal [Internet]. Elsevier Inc.; 2017;20:118–25. Available from: http://dx.doi.org/10.1016/j.jval.2016.09.001

296.Rudarakanchana N, Reeves BC, Bicknell CD, Heatley FM, Cheshire NJ, Powell JT. Editor’s choice - Treatment decisions for descending thoracic aneurysm: Preferences for thoracic endovascular aneurysm repair or surveillance in a discrete choice experiment. Eur J Vasc Endovasc Surg [Internet]. Elsevier Ltd; 2014;48:13–22. Available from: http://dx.doi.org/10.1016/j.ejvs.2014.03.015

297.Bailey K, Cunningham C, Pemberton J, Rimas H, Morrison KM. Understanding Academic Clinicians’ Decision-making for the Treatment of Childhood Obesity. Child Obes [Internet]. 2015;11:696–706. Available from: http://online.liebertpub.com/doi/10.1089/chi.2015.0031

298.Arellano J, González JM, Qian Y, Habib M, Mohamed AF, Gatta F, et al. Physician preferences for bone metastasis drug therapy in Canada. Curr Oncol. 2015;22:e342–8.

299.272. Arellano J, Hauber AB, Mohamed AF, Gonzalez JM, Collins H, Hechmati G, et al. Physicians’ preferences for bone metastases drug therapy in the United States. Value Heal [Internet]. Elsevier; 2015;18:78–83. Available from: http://dx.doi.org/10.1016/j.jval.2014.10.004

300.Oliveira TC, Barlow J, Bayer S. The association between general practitioner participation in joint teleconsultations and rates of referral : a discrete choice experiment. 2015;1–8.

301.Lloyd AJ, Ziani E, Nicolas L, Fordham BA, Soubeyrand B, Bornhöft C. What are the preferences of health care professionals in Germany regarding fully liquid , ready-to-use hexavalent pediatric vaccine versus hexavalent pediatric vaccine that needs reconstitution ? 2015;1517–24.

302.Offerhaus PM, Researcher M, Otten W, Psychologist S, Boxem-tiemessen JCG, Teacher S, et al. Variation in intrapartum referral rates in primary midwifery care in the Netherlands : A discrete choice experiment. Midwifery [Internet]. Elsevier; 2015;31:e69–78. Available from: http://dx.doi.org/10.1016/j.midw.2015.01.005

303.Poulos C, Johnson FR, Krishnarajah G, Anonychuk A, Misurski D. Pediatricians ’ Preferences for Infant Meningococcal Vaccination. Value Heal [Internet]. Elsevier; 2015;18:67–77. Available from: http://dx.doi.org/10.1016/j.jval.2014.10.010

304.Malhotra C, Chan N, Zhou J, Dalager HB, Finkelstein E. Variation in physician recommendations , knowledge and perceived roles regarding provision of end-of-life care. BMC Palliat Care [Internet]. BMC Palliative Care; 2015;1–8. Available from: http://dx.doi.org/10.1186/s12904-015-0050-y

305.Baji P, Gulácsi L, Lovász BD, Golovics PA, Brodszky V, Péntek M, et al. Treatment preferences of originator versus biosimilar drugs in Crohns disease; Discrete choice experiment among gastroenterologists.

Scand J Gastroenterol [Internet]. Informa Healthcare; 2016;51:22–7. Available from: http://dx.doi.org/10.3109/00365521.2015.1054422

306. Blumenschein P, Lilley M, Bakal JA, Christian S. Evaluating stakeholder's perspective on referred out genetic testing in Canada: A discrete choice experiment. Clin Genet. 2016;89:133–8.

307. Fransen J, Kievit W, Neogi T, Schumacher R, Jansen TIM, Dalbeth N, et al. What Should Be the Cut Point for Classification Criteria of Studies in Gout ? A Conjoint Analysis. 2016;68:1731–5.

308. Baji P, Garcı M, Paolucci F. Comparative analysis of decision maker preferences for equity / efficiency attributes in reimbursement decisions in three European countries. 2016;791–9.

309. Baji P, Gulácsi L, Lovász BD, Golovics PA, Péntek M, Rencz F, et al. Treatment preferences of originator versus biosimilar drugs in Crohn ' s disease. Scand J Gastroenterol [Internet]. Informa Healthcare; 2015;5521:0.

310. Morgan JL, Walters SJ, Collins K, Robinson TG, Cheung K, Audisio R, et al. What influences healthcare professionals ' treatment preferences for older women with operable breast cancer ? An application of the discrete choice experiment. Eur J Surg Oncol [Internet]. Elsevier Ltd; 2017;43:1282–7. Available from: http://dx.doi.org/10.1016/j.ejso.2017.01.012

311. Lee H, Bae E. Eliciting preferences for medical devices in South Korea : A discrete choice experiment. Health Policy (New York) [Internet]. Elsevier Ireland Ltd; 2017;121:243–9. Available from: http://dx.doi.org/10.1016/j.healthpol.2017.01.002

312. Bøgelund M, Hagelund L, Asmussen MB. COPD-treating nurses ' preferences for inhaler attributes – a discrete choice experiment. 2017;7995.

313. Hifinger M, Hiligsmann M, Ramiro S, et al Economic considerations and patients' preferences affect treatment selection for patients with rheumatoid arthritis: a discrete choice experiment among European rheumatologists Annals of the Rheumatic Diseases 2017;76:126-132.

314. Munger MA, Walsh M, Godin J, Feehan M. Pharmacist ' s Demand for Optimal Primary Care Service Delivery in a Community Pharmacy : The OPTiPharm Study. 2017; Available from: https://doi.org/10.1177/1060028017722795

315. Boeri M, Longo A, Grisolía JM, Hutchinson WG, Kee F. The role of regret minimisation in lifestyle choices affecting the risk of coronary heart disease. J Health Econ [Internet]. Elsevier B.V.; 2013;32:253–60. Available from: http://dx.doi.org/10.1016/j.jhealeco.2012.10.007

316. Franken M, Koolman X. Health system goals: A discrete choice experiment to obtain societal valuations. Health Policy (New York) [Internet]. Elsevier Ireland Ltd; 2013;112:28–34. Available from: http://dx.doi.org/10.1016/j.healthpol.2012.12.013

317. Hurley J, Mentzakis E. Health-related externalities: Evidence from a choice experiment. J Health Econ [Internet]. Elsevier B.V.; 2013;32:671–81. Available from: http://dx.doi.org/10.1016/j.jhealeco.2013.03.005

318. Laver K, Ratcliffe J, George S, Lester L, Crotty M. Preferences for rehabilitation service delivery: A comparison of the views of patients, occupational therapists and other rehabilitation clinicians using a discrete choice experiment. Aust Occup Ther J. 2013;60:93–100.

319. Ng V, Sargeant JM. A Quantitative Approach to the Prioritization of Zoonotic Diseases in North America: A Health Professionals' Perspective. PLoS One. 2013;8.

320. Rischatsch M, Zweifel P. What do physicians dislike about managed care? Evidence from a choice experiment. Eur J Heal Econ. 2013;14:601–13.

321. Burnett HF, Ungar WJ, Regier DA, Feldman BM, Miller FA. Parents' willingness to pay for biologic treatments in juvenile idiopathic arthritis. Value Heal [Internet]. Elsevier; 2014;17:830–7. Available from: http://dx.doi.org/10.1016/j.jval.2014.08.2668

322. Burton CR, Fargher E, Plumpton C, Roberts GW, Owen H, Roberts E. Investigating preferences for support with life after stroke: a discrete choice experiment. BMC Health Serv Res [Internet]. 2014;14:63. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L601480208%5Cnhttp://dx.doi.org/10.1186/1472-6963-14-63

323. Chu LW, So JC, Wong LC, Luk JK, Chiu PK, Chan CS, et al. Community end-of-life care among Chinese older adults living in nursing homes. Geriatr Gerontol Int. 2014;14:273–84.

324. Cunningham CE, Walker JR, Eastwood JD, Westra H, Rimas H, Chen Y, et al. Modeling mental health information preferences during the early adult years: A discrete choice conjoint experiment. J Health Commun. 2014;19:413–40.

325. Deal K, Keshavjee K, Troyan S, Kyba R, Holbrook AM. Physician and patient willingness to pay for electronic cardiovascular disease management. Int J Med Inform [Internet]. Elsevier Ireland Ltd; 2014;83:517–28. Available from: http://dx.doi.org/10.1016/j.ijmedinf.2014.04.007

326. Determann D, Korfage IJ, Lambooij MS, Bliemer M, Richardus JH, Steyerberg EW, et al. Acceptance of vaccinations in pandemic outbreaks: A discrete choice experiment. PLoS One. 2014;9.

327. Groothuis-Oudshoorn CG, Fermont JM, Van Til JA, Ijzerman MJ. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. BMC Med Inform Decis Mak. 2014;14.

328. Hoefman RJ, Van Exel J, Rose JM, Van De Wetering EJ, Brouwer WBF. A discrete choice experiment to obtain a tariff for valuing informal care situations measured with the CarerQol instrument. Med Decis Mak [Internet]. 2014;34:84–96. Available from: https://doi.org/10.1177/0272989X13492013

329. Huppelschoten AG, Verkerk EW, Appleby J, Groenewoud H, Adang EMM, Nelen WLDM, et al. The monetary value of patient-centred care: Results from a discrete choice experiment in Dutch fertility care. Hum Reprod. 2014;29:1712–20.

330. Pechey R, Burge P, Mentzakis E, Suhrcke M, Marteau TM. Public acceptability of population-level interventions to reduce alcohol consumption: A discrete choice experiment. Soc Sci Med [Internet]. Elsevier Ltd; 2014;113:104–9. Available from: http://dx.doi.org/10.1016/j.socscimed.2014.05.010

331. Plumb AA, Boone D, Fitzke H, Helbren E, Mallett S, Wagner C von, et al. Detection of extracolonic pathologic findings with CT colonography: A discrete choice experiment of perceived benefits. Eur Radiol. 2014;273:144–52.

332. Shono A, Kondo M. Parents' preferences for seasonal influenza vaccine for their children in Japan. Vaccine [Internet]. Elsevier Ltd; 2014;32:5071–6. Available from: http://dx.doi.org/10.1016/j.vaccine.2014.07.002

333. Struik MHL, Koster K, Veldwijk J, Schuit AJ, Lambooij MS. Preference Structure of Clinicians in the Use of Electronic Medical Records; Quantifying the Relative Importance of Barriers and Facilitators of an Innovation. Value Heal [Internet]. 2013;16:A471. Available from: http://www.sciencedirect.com/science/article/pii/S1098301513027629

334. Veldwijk J, Lambooij MS, Bruijning-Verhagen PCJ, Smit HA, Wit GA de. Parental preferences for rotavirus vaccination in young children: A discrete choice experiment. Vaccine [Internet]. Elsevier Ltd;

2014;32:6277–83. Available from: http://dx.doi.org/10.1016/j.vaccine.2014.09.004

335. De Bekker-Grob EW, Bergstra AD, Bliemer MCJ, Trijssenaar-Buhre IJM, Burdorf A. Protective behaviour of citizens to transport accidents involving hazardous materials: A discrete choice experiment applied to populated areas nearby waterways. PLoS One [Internet]. 2015;10:1–14. Available from: http://dx.doi.org/10.1371/journal.pone.0142507

336. Ho MP, Gonzalez JM, Lerner HP, Neuland CY, Whang JM, McMurry-Heath M, et al. Incorporating patient-preference evidence into regulatory decision-making. Surg Endosc Other Interv Tech [Internet]. Springer US; 2015;29:2984–93. Available from: http://dx.doi.org/10.1007/s00464-014-4044-2

337. Chen TT, Tung TH, Hsueh YS, Tsai MH, Liang HM, Li KL, et al. Measuring Preferences for a Diabetes Pay-for-Performance for Patient (P4P4P) Program using a Discrete Choice Experiment. Value Heal [Internet]. Elsevier; 2015;18:578–86. Available from: http://dx.doi.org/10.1016/j.jval.2015.03.1793

338. Kessels R, Herck P Van, Dancet E, Annemans L, Sermeus W. How to reform western care payment systems according to physicians , policy makers , healthcare executives and researchers : a discrete choice experiment. BMC Health Serv Res [Internet]. BMC Health Services Research; 2015;1–14. Available from: http://dx.doi.org/10.1186/s12913-015-0847-7

339. Kjaer NK, Halling A, Pedersen LB. Education for Primary Care General practitioners ' preferences for future continuous professional development : evidence from a Danish discrete choice experiment General practitioners ' preferences for future continuous professional development : evidence. 2015;9879.

340. Lagarde M, Erens B, Mays N. Determinants of the choice of GP practice registration in England : Evidence from a discrete choice experiment. Health Policy (New York) [Internet]. Elsevier Ireland Ltd; 2015;119:427–36. Available from: http://dx.doi.org/10.1016/j.healthpol.2014.10.008

341. Morgan H, Hoddinott P, Thomson G, Crossland N, Farrar S, Yi D, et al. Benefits of Incentives for Breastfeeding and Smoking cessation in pregnancy (BIBS): a mixed-methods study. 2015;19.

342. Powell G, Holmes EAF, Plumpton CO, Ring A, Baker GA, Jacoby A, et al. Pharmacogenetic testing prior to carbamazepine treatment of epilepsy : patients ' and physicians ' preferences for testing and service delivery. 2015;

343. Rischatsch M. Who joins the network ? Physicians ' resistance to take budgetary. J Health Econ [Internet]. Elsevier B.V.; 2015;40:109–21. Available from: http://dx.doi.org/10.1016/j.jhealeco.2014.12.002

344. Landfeldt E, Eriksson J, Ireland S, Musingarimi P, Jackson C, Tweats E, et al. Patient , physician , and general population preferences for treatment characteristics in relapsed or refractory chronic lymphocytic leukemia : A conjoint analysis. Elsevier Ltd; 2016;40:17–23.

345. Makkar SR, Turner T, Williamson A, Louviere J, Redman S, Haynes A, et al. The development of ORACLe : a measure of an organisation ' s capacity to engage in evidence-informed health policy. Heal Res Policy Syst [Internet]. Health Research Policy and Systems; 2016; Available from: http://dx.doi.org/10.1186/s12961-015-0069-9

346. Obse A, Ryan M, Heidenreich S, Normand C, Hailemariam D. Eliciting preferences for social health insurance in Ethiopia : a discrete choice experiment. 2018;1423–32.

347. Goto R, Kakihara H. A discrete choice experiment studying students ' preferences for scholarships to private medical schools in Japan. Hum Resour Health [Internet]. Human Resources for Health; 2016;1–10. Available from: http://dx.doi.org/10.1186/s12960-016-0102-2

348. Mühlbacher A, Johnson FR, Yang J, Happich M, Belger M. Do You Want to Hear the Bad News ? The Value of Diagnostic Tests for Alzheimer ' s Disease. Value Heal [Internet]. Elsevier; 2016;19:66–74.

Available from: http://dx.doi.org/10.1016/j.jval.2015.10.011

349. Pesko MF, Kenkel DS, Wang H, Hughes JM. The effect of potential electronic nicotine delivery system regulations on nicotine product selection. Addiction [Internet]. 2016 [cited 2018 Jun 19];111:734–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26639526

350. Flynn TN, Bilger M, Malhotra C, Finkelstein EA, Flynn TN. Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents ? A Case Study Eliciting Preferences for End-of-Life Care. Pharmacoeconomics. Springer International Publishing; 2016;34:273–84.

351. Wymbs FA, Cunningham CE, Chen Y, Rimas HM, Deal K, Waschbusch DA, et al. Examining Parents ' Preferences for Group and Individual Parent Training for Children with ADHD Symptoms. 2015;45.

352. Wetering EJ, Exel NJA, Rose JM, Hoefman RJ, Brouwer WBF. Are some QALYs more equal than others ? Eur J Heal Econ [Internet]. Springer Berlin Heidelberg; 2016;117–27. Available from: http://dx.doi.org/10.1007/s10198-014-0657-6

353. Kaambwa B, Ratcliffe J, Shulver W, Killington M, Taylor A, Crotty M, et al. Investigating the preferences of older people for telehealth as a new model of health care service delivery : A discrete choice experiment. 2016;0:1–13.

354. Shono A, Kondo M. Mothers ' preferences regarding new combination vaccines for their children in Japan , 2014. Hum Vaccin Immunother [Internet]. Taylor & Francis; 2017;13:766–71. Available from: http://dx.doi.org/10.1080/21645515.2016.1255387

355. Lloyd AJ, Gallop K, Ali S, Hughes D, Lloyd AJ, Gallop K, et al. Social preference weights for treatments in Fabry disease in the UK : a discrete choice experiment. 2017;7995.

356. Gonzalez JM, Johnson FR, Mcateer H, Posner J, Mughal F. Comparing preferences for outcomes of psoriasis treatments among patients and dermatologists in the U . K .: results from a discrete-choice experiment *. 2017;777–85.

357. Schoot T Van De, Pavlova M, Atanasova E, Groot W. Preferences of Bulgarian consumers for quality , access and price attributes of healthcare services — result of a discrete choice experiment. Int J Health Plann Manage. 2017;47–71.

358. Naunheim MR, Rathi VK, Naunheim ML, Alkire BC, Lam AC, Song PC, et al. What Do Patients Want from Otolaryngologists? A Discrete Choice Experiment. 2017; Available from: https://doi.org/10.1177/0194599817717662

359. Jin J, Wang W, He R, Gong H. Valuing health risk in agriculture : a choice experiment approach to pesticide use in China. Environmental Science and Pollution Research; 2017;17526–33.

360. Decalf VH, Huion AMJ, Benoit DF, Petrovic M, Everaert KCMM. Older People ' s Preferences for Side Effects Associated with Antimuscarinic Treatments of Overactive Bladder : A Discrete-Choice Experiment. Drugs Aging. Springer International Publishing; 2017;34:615–23.

361. Deuchert E, Kauer L, Meisen Zannol F., Deuchert E. KL. MZF. Would you train me with my mental. J Ment Health Policy Econ. 2013;16:67–80.

362. Joy SM, Blauvelt BM, Tuncer MA, Beaugrand M, Sangro B, Colombo M, et al. Stakeholder perspectives on European priorities for comprehensive liver cancer control: A conjoint analysis. Eur J Public Health. 2013;23:951–7.

363. Shingler SL, Swinburn P, Ali S, Perard R, Lloyd AJ. A discrete choice experiment to determine patient preferences for injection devices in multiple sclerosis. J Med Econ [Internet]. 2013;16:1036–42. Available

from: http://www.tandfonline.com/doi/full/10.3111/13696998.2013.811079

364. Brown DS, Poulos C, Johnson FR, Chamiec-Case L, Messonnier ML. Preference Measurement in Health. Adv Health Econ Health Serv Res [Internet]. 2014;24:93–121. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84908326857&partnerID=tZOtx3y1

365. Louviere JJ, Flynn TN, Marley AAJ. Best-Worst Scaling [Internet]. Best-Worst Scaling Theory, Methods Appl. 2015 [cited 2018 Jun 22]. Available from: http://ebooks.cambridge.org/ref/id/CBO9781107337855

366. Flynn TN. Valuing citizen and patient preferences in health: Recent developments in three types of best-worst scaling. Expert Rev Pharmacoeconomics Outcomes Res [Internet]. 2010 [cited 2018 Jun 18];10:259–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20545591

367. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. J Health Econ [Internet]. 2007 [cited 2018 Jun 22];26:171–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16707175

Part II

Insights into DCE and BWS-2 challenges and opportunities regarding design and analysis

Chapter 4

Predicting choice shares with discrete choice experiments: using the right model and conducting the right analysis

Donkers B, Soekhai V, Levitan, B, de Bekker-Grob EW

Submitted

Abstract

Preference heterogeneity reflects differences in individuals' preferences within a population. While the importance of preference heterogeneity on patient preferences has been widely discussed, this heterogeneity has additional lesser known implications for the statistical analysis of preference data. For stated preference data in the form of choice data (e.g., discrete choice experiment studies), the multinomial logit model (MNL) does not account for preference heterogeneity while the mixed logit model (MXL) accounts for heterogeneity by estimating the distribution of preferences in the population. Both models can be used to obtain estimates of preference weights, but in many cases more relevant measures like choice shares are of interest. Simulation results reveal that predicted choice shares are also sensitive to heterogeneity and the modeling approach used. When predicting choice shares based only on the estimated population mean of the preference weights, predicted choice shares are more extreme (i.e., less similar than an equal division). Moreover, a method that does not account for preference heterogeneity at all likely outperforms a more complicated statistical model when heterogeneity is ignored in the prediction stages. Using simulations that match an existing study, the error in the choice shares goes down from 11 %-points when going from the right model (MXL) with the mean-based predictions, to 3-4 %-points when using the wrong model (MNL) combined with the corresponding predictions, to 2 %-points for the right model (MXL) with predictions that also account for preference heterogeneity. These results show that conducting the non-corresponding analyses hugely impacts outcomes and subsequent decision-making.

4.1 Introduction

There is an increasing focus on patient preferences in clinical practice guidelines, academic research, and regulatory decision-making.[1–5] Various stakeholders in healthcare agree that not aligning interventions with patient preferences reduces patient adherence.[6] Other arguments for focusing on patient preferences are (i) its increase in patient satisfaction, (ii) decision-making will be more informed and more transparent with the inclusion of patient-relevant value judgments, and (iii) the ethical imperative of accounting for the voice of those using the treatments.[7–11]

A common, important feature of patient preferences is preference heterogeneity, the differences in individuals' preferences within a population. While the importance of preference heterogeneity has been widely discussed, this heterogeneity has additional lesser known implications for the statistical analysis of preference data.[12,13]

Preference heterogeneity will manifest regardless of the preference elicitation method used, including the common methods of discrete choice experiments (DCE) and best-worst-scaling (BWS).[14,15] For both these methods, researchers in healthcare commonly use multinomial logit (MNL) models or mixed logit (MXL) models to infer preference weights, predicted choice shares and other metrics.[16] However, the presence of preference heterogeneity has strong implications for the statistical analysis. A key distinction between MNL and MXL is that MNL does not account for unobserved systematic differences in preferences as it assumes identical preferences for all individuals. The MXL model accounts for preference heterogeneity by estimating the distribution of preferences in the patient population. This modeling flexibility of MXL may therefore yield better support for healthcare decision-making. MNL is however often used to analyze choice data since MXL demands much larger computation times and sample sizes.[17–19] Currently, it is unknown what impact accounting for preference heterogeneity in stated choice modeling has on choice share predictions and therefore on healthcare decision-making. This is particularly critical for assessments that center on patients' decisions to undergo a treatment or to adhere to medication.

The purpose of this paper is threefold. First, to assess whether the accuracy of predicted choice shares in DCE depends on the modelling approach being used – MNL and MXL in particular. Second, to assess the impact of accounting for or ignoring preference heterogeneity in the prediction of choice shares while it is accounted for in model estimation. Third, to assess whether the underlying structure of the population's preference heterogeneity impacts the accuracy of choice share predictions.

4.2 Methods

4.2.1 Model estimation and choice share predictions in DCEs

In this section we briefly introduce MNL and MXL modeling approaches for DCE. We then discuss ways in which the resulting model estimates can be used to predict choice shares. We then demonstrate the potential for biased results that arise when choice share predictions are based on only the estimated population mean of the preference weights.

4.2.1.1 Model-based inference for DCE

In a DCE individuals are asked to select their preferred alternative from a set of alternatives in which it is assumed that: (1) alternatives can be described by their attributes, (2) an individual's valuation depends on the levels of these attributes and (3) choices are based on a specific utility function.[20,21] This utility function underpins the Random Utility Theory (RUT) framework, which assumes choices are based on utility maximization.[18,22–25]

Application of the RUT framework for DCEs generally assumes that utility can be partitioned into a systematic part (V) that is driven by individuals' stable preferences, which the analyst can capture, and an unobserved residual component (ε) representing the part of utility that cannot be captured by the analyst (unobserved utility component). Since the analyst only observes choices and not the underlying true utility levels, probabilistic models are used to account for the unobserved utility component.[18] In the situation of choosing between alternatives j and k, the leads the probability of selecting alternative j over alternative k:

$$P(Y = j) = P\left(V_j + \varepsilon_j > V_k + \varepsilon_k\right) \qquad \text{eq. 1}$$

With Y denoting the chosen alternative, V the systematic utility part for an alternative and ε the unobserved utility part. For ease of exposition, we will assume the systematic utility part to be linear in the preference weights, with the systematic utility for alternative j equal to:

$$V_j = \beta' X_j \qquad \text{eq.2}$$

where X_j represents the vector of observed attributes relating to alternative j and β the preference weights for these attributes.[18] The results described below generalize to more flexible model specifications.

By making assumptions on the unobserved residual component (ε) it becomes possible to statistically estimate the coefficients in eq. 2 from the observed choices in a DCE. The MNL model and its generalizations, which are commonly used for the analysis of DCE data[18,22], follows when the error term is independently and identically distributed (IID) and extreme value (EV) type I across alternatives.[18,25] This results in the probability that an individual i prefers alternative j over all other alternatives in S that equals:[8]

$$P(Y_i = j) = \frac{exp\left(V_{ij}\right)}{\sum_{s=1}^{S} exp\left(V_{is}\right)} \qquad \text{eq.3}$$

A limitation of MNL is that all individuals are assumed to have the same preference weights, hence ignoring unobserved, yet systematic differences in preferences (preference heterogeneity).[26] To overcome this limitation, alternative models such as MXL are used. Individual level choice shares in MXL are predicted in the same way as in eq.3, but the systematic utility part is now defined as:

$$V_{ij} = \beta_i{}' X_j \qquad \text{eq.4}$$

with β_i being specific to individual *i*. These individual level preference weights are then assumed to follow a specific distribution. We will assume a normal distribution with mean μ_β and covariance matrix Σ_β, but our results generalize for other distributions.[26] The presence of this distribution of individual level preference weights provides a structure to incorporate preference heterogeneity.[18]

4.2.1.2 Predicting choice shares in DCEs using MNL and MXL estimates

Predicted choice shares for a specific set of alternatives are often used to convey DCE results to policy makers for decision-making.[27,28] Estimates of both MNL and MXL can be used to predict choice shares.

Consider a choice with two alternatives, j and k, characterized by X_j and X_k. When an MNL model is estimated, choice shares will be predicted as:

$$P(Y = j|\beta, X_j, X_k) = \frac{exp(X_j\beta)}{exp(X_j\beta) + exp(X_k\beta)} \qquad \text{eq.5}$$

Here β represents the estimated preference weights from the MNL model. Note that this choice share prediction does not account for preference heterogeneity.

With MXL, there are two approaches to predicting choice shares. The first approach follows the MNL approach and predicts choice shares the estimated mean population preference parameters (μ_β) for each attribute (level). This leads to predicted choice shares comparable to the MNL-based predictions in eq.5 with β being replaced by μ_β. However, this approach ignores the heterogeneity in preferences that was modelled. Alternatively, one can also account for preference heterogeneity in the choice share predictions. Here, the prediction in eq. 5 is made for each possible β_i, and the final choice share prediction takes the expected value of the choice shares given β_i, with respect to its estimated distribution. The expected choice share for DCE options j and k is then given by:

$$P(Y = j|X_j, X_k, \mu_\beta, \Sigma_\beta) = \int \frac{exp(X_j\beta_i)}{exp(X_j\beta_i) + exp(X_k\beta_i)} f(\beta_i|\mu_\beta, \Sigma_\beta)d\beta_i \qquad \text{eq.6}$$

Here $f(\beta_i|\mu_\beta, \Sigma_\beta)$ represents the probability density of the distribution of the individual level preference weights, in our case the density of the $N(\mu_\beta, \Sigma_\beta)$ distribution. In practice, the integral can be approximated using a sample of preference parameters drawn from the specified distribution. Using this procedure, preference heterogeneity is taken into account not only during model estimation but also when predicting choice shares. The resulting choice share predictions are based on the full distribution of individual-level preferences under MXL.

Contrasting the choice share prediction based on the estimated mean preference coefficients (i.e. eq. 5 with $\beta = \mu_\beta$) with the predictions that account for preference heterogeneity (i.e. eq. 6), we anticipate a bias for the mean-based prediction that drives the predictions away from a 50%/50% choice share (between the two alternatives), i.e. the predictions shift to more extreme choice shares. Our reasoning is as follows: in the presence of preference heterogeneity, the population choice share will be the expected value of the choice share for preference parameters drawn from the population preference distribution. Jensen's inequality states that for a concave (convex) function the expected value of a function is larger (smaller) than the function when evaluated in the expected value of the input.[29] The shape of the logit probability function is such that for values below a 50% predicted probability, the function is concave and hence when evaluated at the mean preference parameters it will be biased towards 0%. For probabilities above 50%, the function is convex and hence predictions will be biased towards 100%.

Our argument above deserves some further remarks. The normal distribution has an unbounded support, so when computing the expected value in eq. 6 both the convex and concave parts of the probability function receive positive weights. The majority of the weight, however, is assigned to the region that biases the predictions away from 50%/50%. For other types of preference heterogeneity distributions, especially asymmetric ones, the net effect is not immediately clear. There could be rare cases where the predictions for a specific set of choice alternatives will be unbiased. In general, however, we anticipate that a bias away from 50%/50% will be present.

4.2.1 Simulation

To assess whether the accuracy of DCE choice share predictions depends in a structured manner on the model being used (MNL vs MXL), the type of choice share prediction and the underlying structure of the population preference heterogeneity, we simulate data based on an existing DCE study. The study from de Bekker-Grob et al. - which focused on patients' preferences for surgical management of esophageal cancer - was used as base case for our simulation, with the same attributes and levels (Table 1).[30] Mean coefficients and standard deviations from the de Bekker-Grob et al. study were used respectively as mean population parameters (μ) and preference heterogeneity (σ) in our simulation.[30] The μ's and σ's for attributes 2-6 were constant across all scenarios (Table 1). Individual level preferences for an attribute level were assumed to be independent of all other preference weights (no correlations).

Table 1 Attributes, levels and parameters of the heterogeneity distribution used in simulation

Attributes	Levels	μ	σ
1. In-hospital mortality	2%/5%/10%	See Table 2	See Table 2
2. Persistent gastrointestinal symptoms	10%/40%/80%	-0.07400	0.03800
3. 5-year survival after esophagectomy	20%/35%/50%	0.23300	0.21400
4. Risk for postoperative complications (morbidity)	20%/40%/60%	-0.06700	0.03800
5. Hospital volume medium (dummy)	0/1	1.11000	0.06700
6. Hospital volume high (dummy)	0/1	1.89000	0.92000

Fifteen simulation scenarios were developed to study whether the accuracy of predicted choice shares depend on the amount of population preference heterogeneity (Table 2). This approach we designed to ensure our results apply for realistic levels of heterogeneity. The scenarios focused on the effect of varying the preference distribution of the first attribute, in-hospital mortality. In simulation scenarios 1-5, increasing values of μ were used with the scenario 3 having the baseline value, i.e. the value corresponding to the de Bekker-Grob et al. study. Heterogeneity was set at the baseline value of 0.146. In scenarios 6-10, population preference heterogeneity was increased by varying the σ's for the first attribute, keeping everything else constant. To investigate whether outcomes were driven by the values of μ and σ separately, or by the ratio σ/μ, scenarios 11-15 varied μ and σ for the first attribute while keeping both σ/μ and all other parameters constant.

To measure the performance of the choice share predictions for each simulation scenario, we used the mean absolute error (MAE). The MAE quantifies the average absolute error of a prediction (comparing predicted value and true value).[31] As the prediction error is likely to be sensitive to the specific set of alternatives for which the choice share prediction is made. To overcome this sensitivity, we compute the MAE by averaging the absolute errors across a total of S=1296 choice sets consisting of all unique pairs of alternatives that had no overlap in attribute levels. This means that for each replication *r* the MAE_r is calculated by averaging over the S choice sets as follows:

$$MAE_r = \frac{1}{S}\sum_{s=1}^{S}\left|\widehat{choice\ shares}_{r,s} - choice\ shares_s\right| \qquad \text{eq.7}$$

Here $choice\ shares_s$ represents the expected population-level choice share for set s and $\widehat{choice\ shares}_{r,s}$ the corresponding predicted choice share for replication r. A lower MAE indicates a better predictive performance.

Table 2 Simulation scenarios with μ and σ values for the attribute in-hospital mortality, including absolute σ/μ value.

	Scenario	μ In-hospital mortality	σ In-hospital mortality	Absolute σ/μ value
↓ σ/μ value	1	**-0.09175**	0.14600	1.59128
	2	**-0.18350**	0.14600	0.79564
	3	**-0.36700**	0.14600	0.39782
	4	**-0.73400**	0.14600	0.19891
	5	**-1.46800**	0.14600	0.09946
↑ σ/μ value	6	-0.36700	**0.03650**	0.09946
	7	-0.36700	**0.07300**	0.19891
	8	-0.36700	**0.14600**	0.39782
	9	-0.36700	**0.29200**	0.79564
	10	-0.36700	**0.58400**	1.59128
constant σ/μ value	11	**-0.09175**	**0.03650**	0.39782
	12	**-0.18350**	**0.07300**	0.39782
	13	**-0.36700**	**0.14600**	0.39782
	14	**-0.73400**	**0.29200**	0.39782
	15	**-1.46800**	**0.58400**	0.39782

All simulation scenarios used a matching Bayesian D-efficient experimental design generated using Ngene software (Ngene, version 1.2.1). These designs with 24 choice tasks each were generated using normally distributed priors that correspond to the actual population distribution (cf. Tables 1 and 2). 1,000 replications of a sample of size 500 were simulated for each scenario. MNL and MXL models were estimated on each sample and choice share predictions obtained.[31] The simulation code was written in Julia programming language version 1.5.0 (https://julialang.org/).

4.3 Results

Figures 1-3 present box plots for the distribution of MAE_r across the 1,000 replications for the 15 simulation scenarios. The squares in each boxplot represent the average MAE across the 1,000 replications. Figure 1 shows results for MNL, Figure 2 shows results for MXL-μ_β (i.e. mean-based, not accounting for preference heterogeneity)and Figure 3 corresponds to MXL-full (i.e. fully accounting for preference heterogeneity).

For MNL, there was variation in MAE values across the 15 simulation scenarios (Figure 1). In scenarios 1-5, where the mean preference weight was increased, the average MAE varied around 4 %-points in scenario 2, to 2.5 %-points in scenario 5. In scenarios 6-10, where the heterogeneity was increased, the average MAE varied between 3 %-points and nearly 5 %-points, while the last set of scenarios 11-15 (constant ratio of σ/μ) showed a relatively stable MAE of about 3.6 %-points. Overall, the average MAE for MNL was 3 to 4 %-points. The width of the boxplots, which represents the reliability of the assessment, does not reveal a clear relationship between the parameters of the preference heterogeneity distribution and the variation in the MAE across replications.

In Figure 2, which presents a box plot of the MAE values for MXL-μ_β, average MAE values were around 11 %-points for most of the simulation scenarios. MAE values for

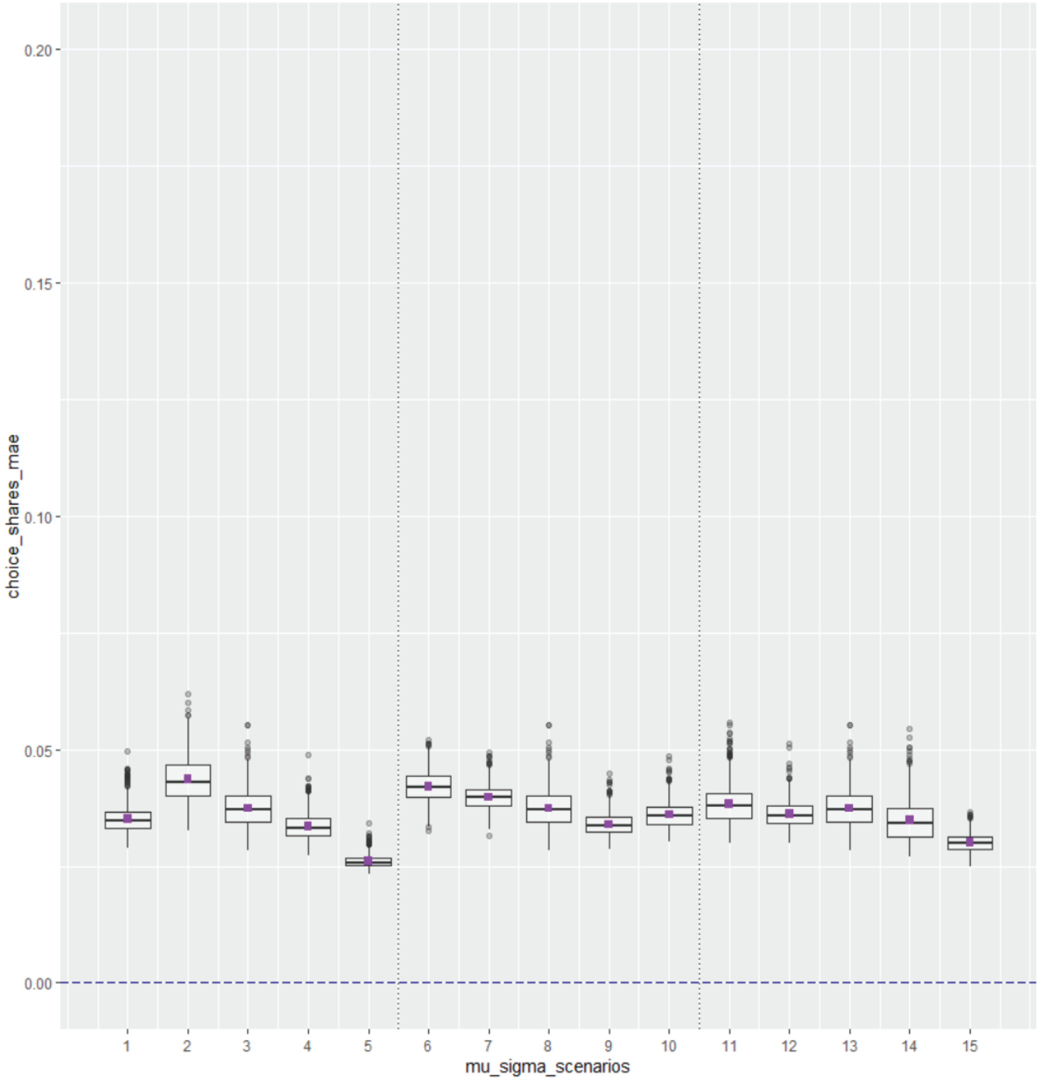


Figure 1 Box plot of MAE for MNL choice shares for 15 simulation scenarios. The purple square indicates the mean MAE.

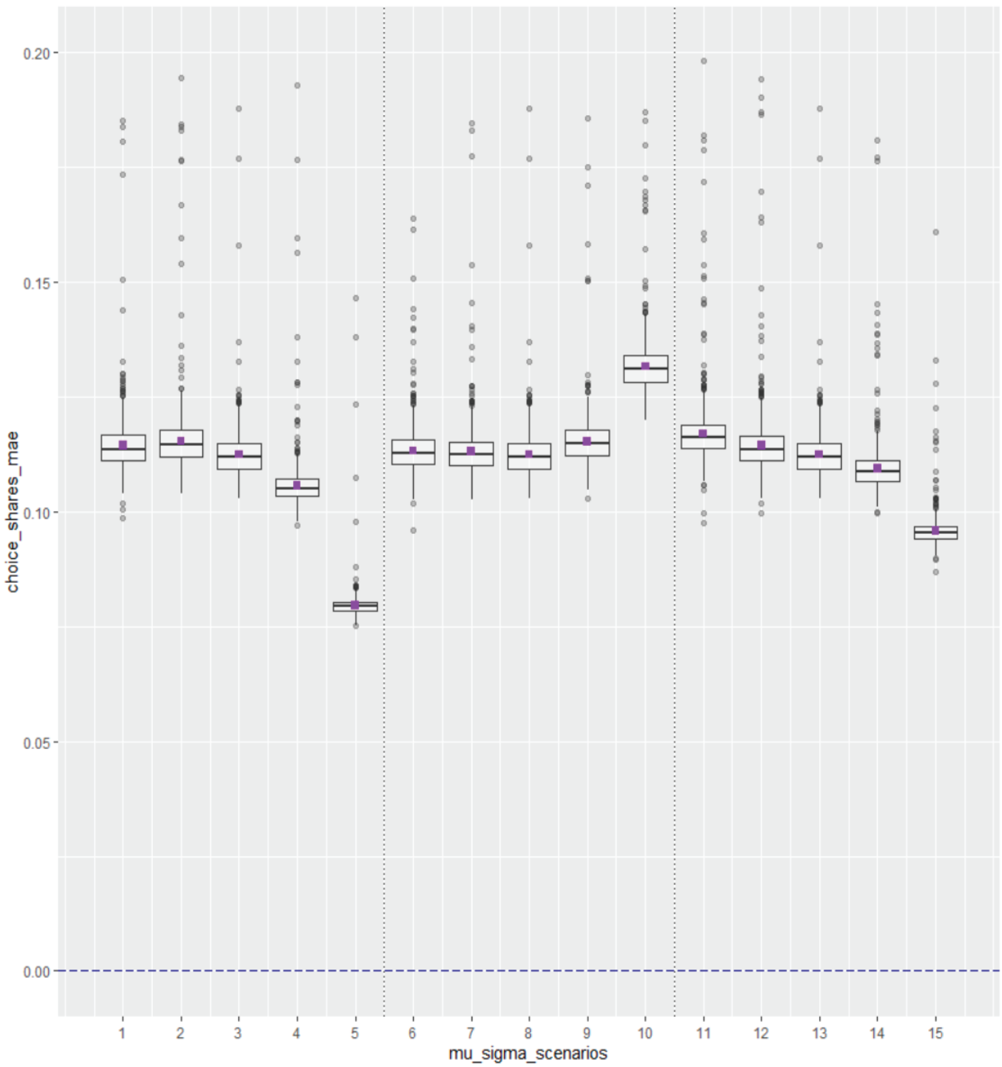


Figure 2 Box plot of MAE for MXL predictions not accounting for preference heterogeneity (MXL-μ_β) for 15 simulation scenarios. The purple square indicates the mean MAE.

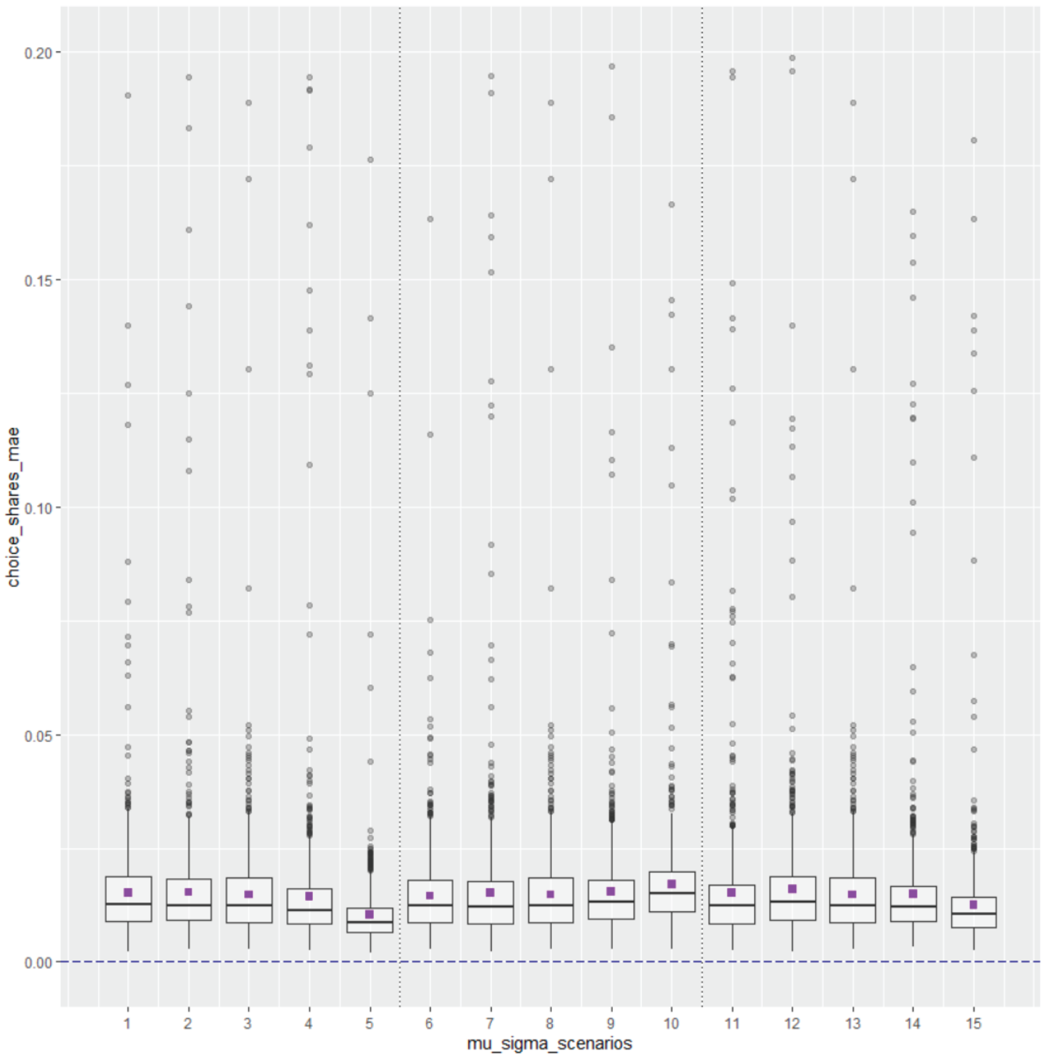


Figure 3 Box plot of MAE for MXL predictions fully accounting for preference heterogeneity (MXL-full) for 15 simulation scenarios. The purple square indicates the mean MAE.

simulation scenarios 5 and 15 were lower, while scenario 10 showed the highest MAE value. Looking at the overall trend, choice share predictions based on estimated population means that ignore preference heterogeneity led to MAE values that were generally larger than 10 %-points. The width of the boxplots became smaller for scenarios 1-5, but not larger for scenarios 6-10 with increasing heterogeneity and were not constant across scenarios 11-15.

Figure 3 presents the box plots of the MAE for MXL-full choice share predictions. It shows that for all simulation scenarios the average MAE value was around 2%-points, with some variation in MAE values across the scenarios. The width of the boxplots showed a similar pattern as for MXL-μ_β.

Looking at the overall trend in choice share MAE values (Table 3), we find MXL taking preference heterogeneity into account by looking at the full distribution of preferences (Figure 3) outperforms the other methods with the lowest average MAE values (around 2%-points). Using MNL to predict choice shares performs second best in this case (Figure 1), with average MAE values typically around 3-4 %-points (Figure 1). Using MXL and ignoring preference heterogeneity (MXL-μ_β) when predicting choice shares performs worst (Figure 2), with average MAE values around 11%-points. It should also be noted that the variation in MAE values increases for all simulation scenarios when using MXL compared to using MNL.

Studying the general patterns across the 15 scenarios for all methods, we saw that the variation in MAE, quantified by the width of the boxplots, did not increase with increasing levels of preference heterogeneity. The boxplots also showed that for the largest heterogeneity and mean preference value (scenario 10), the mean MAE value was the largest for both MXL-μ_β and MXL-full, but not for MNL. The variation in MAE across replications is largest for MXL-full.

Table 3 Mean, median, Q1 and Q3 MAE values for MNL, MXL-μ_β and MXL-full for each simulation scenario

Scenarios	MAE MNL				MAE MXL-μ_β				MAE MXL-full			
	Mean	Median	Q1	Q3	Mean	Median	Q1	Q3	Mean	Median	Q1	Q3
1	0.03516	0.03468	0.03311	0.03668	0.11534	0.11371	0.11113	0.11679	0.01629	0.01280	0.00892	0.01890
2	0.04362	0.04303	0.04010	0.04684	0.11590	0.11470	0.11190	0.11770	0.01574	0.01249	0.00910	0.01823
3	0.03738	0.03702	0.03446	0.03994	0.11240	0.11190	0.10920	0.11480	0.01476	0.01248	0.00861	0.01841
4	0.03353	0.03315	0.03142	0.03524	0.10585	0.10503	0.10336	0.10707	0.01480	0.01141	0.00827	0.01613
5	0.02603	0.02577	0.02501	0.02683	0.07955	0.07935	0.07829	0.08031	0.01029	0.00866	0.00652	0.01191
6	0.04210	0.04204	0.03983	0.04419	0.11331	0.11267	0.11033	0.11558	0.01447	0.01233	0.00873	0.01805
7	0.03979	0.03967	0.03796	0.04141	0.11350	0.11250	0.11010	0.11530	0.01556	0.01218	0.00829	0.01769
8	0.03738	0.03702	0.03446	0.03994	0.11240	0.11190	0.10920	0.11480	0.01476	0.01248	0.00861	0.01841
9	0.03397	0.03372	0.03224	0.03540	0.11590	0.11500	0.11240	0.11770	0.01615	0.01305	0.00935	0.01808
10	0.03606	0.03571	0.03394	0.03768	0.13230	0.13110	0.12820	0.13420	0.01739	0.01495	0.01115	0.01991
11	0.03838	0.03780	0.03539	0.04061	0.11742	0.11618	0.11381	0.11890	0.01582	0.01231	0.00824	0.01695
12	0.03615	0.03577	0.03409	0.03793	0.11482	0.11352	0.11103	0.11662	0.01672	0.01313	0.00927	0.01865
13	0.03738	0.03702	0.03446	0.03994	0.11240	0.11190	0.10920	0.11480	0.01476	0.01248	0.00861	0.01841
14	0.03485	0.03413	0.03138	0.03747	0.10958	0.10867	0.10670	0.11108	0.01500	0.01214	0.00895	0.01660
15	0.03006	0.02991	0.02861	0.03130	0.09585	0.09552	0.09420	0.09687	0.01245	0.01042	0.00753	0.01420

4.4 Discussion

In this paper we studied whether the accuracy of choice share predictions for choice options in a DCE depends on the modelling approach (MNL or MXL), the type of analysis (accounting for preference heterogeneity or not) being used and the underlying structure of the population preference heterogeneity. To quantify how well each approach performed, MAE values were predicted and plotted against each of the simulation scenarios. Our results showed that using

MXL for choice modelling and accounting for preference heterogeneity leads to the lowest MAE, with an average MAE of about 2 %-points. Using the MNL approach (i.e., not accounting for preference heterogeneity at all) led to larger average MAE values (around 3-4 %-points). Finally, MXL-μ_β performs worst with average MAE values around 11 %-points. Based on these outcomes we can state that, assuming that there is evidence for preference heterogeneity in the choice data, the right model with the non-corresponding analysis (MXL-μ_β) performs worse than the wrong model with the wrong but corresponding analysis (MNL).

In DCE literature there is a large amount of work that states that when dealing with preference heterogeneity, MXL is one of the modelling approaches that can be used to overcome the often unrealistic assumption in MNL that individuals have identical preferences.[18,22,26] Textbook DCE modelling approaches suggest that when dealing with preference heterogeneity, MXL will most likely lead to more accurate estimated coefficients compared to MNL since individual preference heterogeneity is taken into account. In this paper, however, we focused on the prediction of choice shares as a more policy-relevant feature that goes beyond the estimated coefficients.

When executed correctly, choice share predictions based on MXL are overall more accurate than those of MNL.[33,34] However, in practice, multiple studies focus on the estimated population mean preference parameters to perform policy simulations and recommendations. Examples include a study from Determann et al. (2014)[35] and Grausman et al. (2021)[36] in which estimated population mean preference parameters were used to predict choice shares. Even though these studies relied on latent-class analyses (LCA), similar to MXL, relying on mean preference parameters by averaging preference weights with respect to the class probabilities before predicting choice shares will lead to biased choice share predictions.

For the particular preference problem we studied, our results show that the MXL mean-based approach for predicting choice shares is worse than the simpler MNL approach, and that the MXL heterogeneity-based approach for predicting choice share works better than both. This means that when dealing with preference heterogeneity, using the wrong model (MNL), but with a corresponding prediction approach, performs on average 7-8%-points better than using MXL-μβ, which relies on the right model (MXL), but with the non-corresponding (mean-based) analysis. In the presence of preference heterogeneity, analysts are hence better off not accounting for preference heterogeneity at all, instead of modelling heterogeneous preferences but not accounting for it when drawing policy implications based on choice share predictions. The best approach, however, is the advanced MXL model with choice share predictions that account for preference heterogeneity, i.e. MXL-full. We note that the specific performance differences only apply for the current application. However, it is likely that our conclusions are also relevant within other contexts, even though it is difficult to predict the precise magnitude of the differences. Our results show that differences in the heterogeneity distribution across the 15 scenarios are relatively minor, but more research is needed to understand how the performance of each method varies with the specifics of the application.

Additionally, one of our results was that we saw that larger heterogeneity did not affect the width of the boxplots. This tells us that in our simulation study increasing heterogeneity did not impact the reliability of our choice shares MAE assessment. Future research should focus on the specific impact of larger heterogeneity on the reliability of choice share assessments.

This study has been limited by the fact that only one case study has been assessed, using the mean coefficients and standard deviations from the study from de Bekker-Grob et al. as inputs for our simulation study (see Tables 1 and 2). This means that the results that were found in this study are based on inputs from one case study only. Using for example inputs that are on the other side of the utility space, could theoretically lead to different outcomes. In order to test the robustness of our outcomes, future research should focus on other inputs than the ones that were used in this study. We do however believe that the differences in choice shares MAE we found in this study provide first evidence that the modelling and analysis approach impacts the accuracy of predicted choice shares in DCE.

4.5 Conclusion

This study demonstrated that predicted choice shares are sensitive to the modelling (i.e., MNL or MXL) and type of analysis approach taken (i.e., accounting for preference heterogeneity or not). For the particular case studied, we showed that using MXL and accounting for preference heterogeneity leads to the lowest average MAE values. Using MXL in estimation, but not accounting for heterogeneity in the prediction of the choice shares resulted in average MAE values that exceed those of the choice share predictions based on MNL. Hence, our results showed that – when predicting choice shares – researchers might be better off ignoring preference heterogeneity than accounting for it during estimation but ignoring it in prediction. This is troubling as most software applications facilitate estimation of models that account for heterogeneity, but only a few support researchers in handling the complexity that arises in prediction. As our results showed that conducting non-corresponding analysis hugely impacts choice share prediction quality and therefore decision-making, researchers will need to account for preference heterogeneity also at the prediction stage of their research. This requires additional skills from researchers or added functionality of available software packages.

References

1. Boivin A, Green J, Van Der Meulen J, Légaré F, Nolte E. Why consider patients' preferences?: A discourse analysis of clinical practice guideline developers. *Med Care*. 2009;47(8):908-915. doi:10.1097/MLR.0b013e3181a81158

2. Domecq JP, Prutsky G, Elraiyah T, et al. Patient engagement in research: A systematic review. *BMC Health Serv Res*. 2014;14(1):1-9. doi:10.1186/1472-6963-14-89

3. Whitty JA, Fraenkel L, Saigal CS, Groothuis-Oudshoorn CGM, Regier DA, Marshall DA. Assessment of Individual Patient Preferences to Inform Clinical Practice. *Patient*. 2017;10(4):519-521. doi:10.1007/s40271-017-0254-8

4. Ho M, Saha A, McCleary KK, et al. A Framework for Incorporating Patient Preferences Regarding Benefits and Risks into Regulatory Assessment of Medical Technologies. *Value Heal*. 2016;19(6):746-750. doi:10.1016/j.jval.2016.02.019

5. Levitan B, Hauber AB, Damiano MG, Jaffe R, Christopher S. The Ball is in Your Court: Agenda for Research to Advance the Science of Patient Preferences in the Regulatory Review of Medical Devices in the United States. *Patient*. 2017;10(5):531-536. doi:10.1007/s40271-017-0272-6

6. Ostermann J, Brown DS, de Bekker-Grob EW, Mühlbacher AC, Reed SD. Preferences for Health Interventions: Improving Uptake, Adherence, and Efficiency. *Patient*. 2017;10(4):511-514. doi:10.1007/s40271-017-0251-y

7. Huls SPI, Whichello CL, van Exel J, Uyl-de Groot CA, de Bekker-Grob EW. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. *Value Heal*. 2019;22(11):1318-1328. doi:10.1016/j.jval.2019.04.1930

8. Scott AM, Wale JL. Patient advocate perspectives on involvement in HTA: An international snapshot. *Res Involv Engagem*. 2017;3(1). doi:10.1186/s40900-016-0052-9

9. Dirksen CD. The use of research evidence on patient preferences in health care decision-making: Issues, controversies and moving forward. *Expert Rev Pharmacoeconomics Outcomes Res*. 2014;14(6):785-794. doi:10.1586/14737167.2014.948852

10. Bridges JFP, Jones C. Patient-based health technology assessment: A vision of the future. *Int J Technol Assess Health Care*. 2007;23(1):30-35. doi:10.1017/S0266462307051549

11. Hansen HP, Lee A. Patient aspects and involvement in HTA: An academic perspective. *Pharm Policy Law*. 2011;13(3-4):123-128. doi:10.3233/PPL-2011-0319

12. Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *J Health Econ*. 2008;27(4):1078-1094. doi:10.1016/j.jhealeco.2007.11.006

13. Hess S, Meads D, Twiddy M, Mason S, Czoski-Murray C, Minton J. Characterising heterogeneity and the role of attitudes in patient preferences: A case study in preferences for outpatient parenteral intravenous antimicrobial therapy (OPAT) services. *J Choice Model*. 2021;38:100252. doi:10.1016/j.jocm.2020.100252

14. Soekhai V, Whichello C, Levitan B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324-1331. doi:10.1016/j.drudis.2019.05.001

15. Whichello C, Levitan B, Juhaeri J, et al. Appraising patient preference methods for decision-making in the medical product lifecycle: An empirical comparison. *BMC Med Inform Decis Mak*. 2020;20(1):1-15. doi:10.1186/s12911-020-01142-w

16. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics*. 2019;37(2):201-226. doi:10.1007/s40273-018-0734-2

17. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *Patient*. 2015;8(5):373-384. doi:10.1007/s40271-015-0118-z

18. Train K. *Discrete Choice Methods with Simulation*.; 2002.

19. Hess S, Train KE. Recovery of inter- and intra-personal heterogeneity using mixed logit models. *Transp Res Part B Methodol*. 2011;45(7):973-990. doi:10.1016/j.trb.2011.05.002

20. Lancaster KJ. A New Approach to Consumer Theory. *J Polit Econ*. 1966;74(2):132-157. doi:10.1086/259131

21. Thurstone LL. A law of comparative judgment. *Psychol Rev*. 1927;34(4):273-286. doi:10.1037/h0070288

22. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*.

23. Ben-Akiva ME, Lerman SR. *Discrete Choice Analysis : Theory and Application to Travel Demand*. MIT Press; 1985.

24. Mcfadden D. *Economic Choices*. Vol 91.; 2001.

25. McFadden D. Conditional logit analysis of qualitative choice behavior. *Front Econom*. Published online 1974.

26. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Heal*. 2016;19(4):300-315. doi:10.1016/j.jval.2016.04.004

27. Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc*. 2015;29(10):2984-2993. doi:10.1007/s00464-014-4044-2

28. de Bekker-Grob EW, Veldwijk J, Jonker M, et al. The impact of vaccination and patient characteristics on influenza vaccination uptake of elderly people: A discrete choice experiment. *Vaccine*. 2018;36(11):1467-1476. doi:10.1016/j.vaccine.2018.01.054

29. Simic S. Jensen’s inequality and new entropy bounds. *Applied Mathematics Letters*. 2009;22(8):1262-1265.

30. De Bekker-Grob EW, Niers EJ, Van Lanschot JJB, Steyerberg EW, Wijnhoven BPL. Patients’ Preferences for Surgical Management of Esophageal Cancer: A Discrete Choice Experiment. *World J Surg*. 2015;39(10):2492-2499. doi:10.1007/s00268-015-3148-8

31. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci Model Dev*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014

32. Koehler E, Brown E, Haneuse SJPA. On the assessment of Monte Carlo error in simulation-based Statistical analyses. *Am Stat*. 2009;63(2):155-162. doi:10.1198/tast.2009.0030

33. Hess S, Polak JW. Mixed logit modelling of airport choice in multi-airport regions. *J Air Transp Manag*. 2005;11(2):59-68. doi:10.1016/j.jairtraman.2004.09.001

34. Alfnes F. Stated preferences for imported and hormone-treated beef: Application of a mixed logit model. *Eur Rev Agric Econ*. 2004;31(1):19-37. doi:10.1093/erae/31.1.19

35. Determann D, Korfage IJ, Lambooij MS, et al. Acceptance of vaccinations in pandemic outbreaks: A discrete choice experiment. *PLoS One*. 2014;9(7):e102505. doi:10.1371/journal.pone.0102505

36. Grauman Å, Hansson · Mats, James S, Hauber · Brett, Veldwijk J. The Patient-Patient-Centered Outcomes Research Communicating Test Results from a General Health Check: Preferences from a Discrete Choice Experiment Survey. *Patient - Patient-Centered Outcomes Res*. Published online 123AD. doi:10.1007/s40271-021-00512-4

Chapter 5

Best Worst Scaling: for Good or for Bad but not for Both

Soekhai V, Donkers B, Levitan, B, de Bekker-Grob EW

Published Journal of Choice Modelling 2021

Abstract

This paper studies the performance of case 2 best-worst scaling (BWS) when it is applied to a mix of positive and negative attributes, for example in studying treatments characterized by both benefits and harms. Intuitively, such a mix of positive and negative attributes leads to dominance. We analytically show that dominance leads to infinitely large differences between the parameter estimates for the positive versus negative attributes. The results from a simulation study confirm our analytical results: parameter values of the attributes could not be accurately recovered. When only a single positive attribute was used, even the relative ordering of the attribute level preferences was not identified. As a result, case 2 BWS can be used to elicit preferences if only good (positive) or only bad (negative) attributes are included in the choice tasks, but not for both since dominance will impact parameter estimation and therefore decision-making.

5.1 Introduction

Best-worst scaling (BWS) has become an increasingly popular method to elicit preferences in health and healthcare (Flynn et al., 2007; Mühlbacher et al., 2016a). The introduction of BWS came from the intent to obtain more preference information than from a discrete choice experiment (DCE) by asking individuals to select their "best" and "worst" option, without increasing the cognitive burden (Louviere et al., 2015; Thurstone, 1927). BWS in health economics is commonly used for health state valuation and medical treatment valuation (Mühlbacher et al., 2016a). However, there are also many other areas of BWS applications; e.g. for health policy making, patient and expert preference assessment and benefit-risk assessment (Hollin et al., 2017; Mühlbacher et al., 2016b; Severin et al., 2013 Tarini et al., 2018).

In BWS literature it is stated that BWS is a more efficient way to elicit preferences compared to a "pick one" task, therefore providing more information, since individuals are asked to select their "best" and "worst" option (Flynn et al., 2007). There are three types of BWS: object case (case 1 BWS), profile case (case 2 BWS) and the multi-profile case (case 3 BWS) (Louviere et al., 2015).[v] The object case (Figure 1a) shows several attributes from which individuals choose the attributes they consider "best" (or for example "most important") and "worst" ("least important"). The profile case (Figure 1b) looks similar to the object case but differs in that it presents individuals levels of attributes which form a so-called 'profile' (e.g. the attributes of a medical treatment), and individuals explicitly value the levels of attributes instead of the attributes themselves (Flynn et al., 2007) by making "best" and "worst" choices. Compared to the profile case, the multi-profile case (Figure 1c) includes two or more profiles where individuals choose their "best" and "worst" profiles. The multi-profile

[v] Different terminology is sometimes used in other disciplines when referring to BWS types.

case is similar to a regular DCE, except that the BWS type also includes a “worst” choice, which is not the case in a traditional DCE.

a

Best		Worst
[ ]	Being cured	[ ]
[ ]	Severe side effects	[ ]
[ ]	Voice changes	[ ]
[ ]	Calcium deficiency	[ ]

b

Best		Worst
[ ]	Chance of being cured: 40%	[ ]
[ ]	Chance of severe side effects: 2%	[ ]
[ ]	Chance of voice changes: 5%	[ ]
[ ]	Chance of calcium deficiency: 5%	[ ]

c

Treatment A	Treatment B	Treatment C	
Chance of being cured: 40%	Chance of being cured: 70%	Chance of being cured: 100%	
Chance of severe side effects: 2%	Chance of severe side effects: 10%	Chance of severe side effects: 5%	
Chance of voice changes: 5%	Chance of voice changes: 0%	Chance of voice changes: 10%	
Chance of calcium deficiency: 5%	Chance of calcium deficiency: 10%	Chance of calcium deficiency: 15%	
[ ]	[ ]	[ ]	Best
[ ]	[ ]	[ ]	Worst

Figure 1 Examples of the three BWS cases, with a. case 1 BWS (object case), b. case 2 BWS (profile case) and c. case 3 BWS (multi-profile case)

Case 2 BWS experiments especially received much attention in health economics, as they can uncover attribute level importance, reduce cognitive burden of the elicitation task by focusing on one profile at a time and are relatively easy to design (van Dijk et al., 2016; Whitty et al., 2014). While much is already known about case 3 BWS due to its similarities to

DCEs, case 2 BWS is still in its infancy and several issues relating to its design and analysis require further exposition. One of these issues is the inclusion of a mixture of positive (e.g. benefit) and negative (e.g. harm) attributes. In this paper we will show that case 2 BWS with such a mixture of attributes will lead to estimation problems through the concept of dominance.

Within DCEs, there is a considerable amount of work about the impact of dominance on parameter estimation and evidence suggests that it can significantly bias the estimated parameters (Bliemer and Rose, 2011; Flynn et al., 2008; Huber et al., 1982; Tervonen et al., 2018). Although a study by Krucien et al. (2017) suggests the relevance of investigating the impact of dominant attributes in BWS and Flynn (2010) hinted towards potential estimation problems when dealing with dominant attributes in BWS, there is little research about the specific impact of dominance in BWS experiments on parameter estimates. Obtaining insights into this topic is important, especially for case 2 BWS due to its increased popularity in health economic research.

The aim of this paper is to study the effect of a mixture of positive and negative attributes on parameter estimates in case 2 BWS experiments. This will be illustrated both analytically and with simulation examples. This study will be an important step to further advance our understanding of case 2 BWS experiments.

5.2 Dominant attributes in case 2 BWS

In this section we elaborate on dominant attributes in case 2 BWS, including the choice process in case 2 BWS with dominant attributes.

5.2.1 Dominant attributes

In this paper we define a dominant attribute analogous to the definition of dominant alternatives in discrete choice experiments (DCE) (Bliemer et al., 2017; Bliemer and Rose, 2011; Huber et al., 1982): a dominant attribute is the attribute that is always selected as "best" (or "worst") since all its levels are preferred over all levels of every other attribute. Individuals in case 2 BWS select "best" and "worst" attribute levels and not attributes (Louviere et al., 2015).

In this paper we will show how attribute dominance arises and how it affects model estimation. Building on theories of behavioral economics (Levin and Gaeth, 1988; Tversky and Kahneman, 1981), we define positive attributes as attributes generally interpreted as a "gain" (e.g. increased life expectancy or increased probability of getting cured). Similarly, negative attributes are defined as attributes generally interpreted as a "loss" (e.g. increased treatment costs or increased risk of heart failure). In general, people prefer gains over losses (Kahneman and Tversky, 1988). This means that a case 2 BWS experiment with one positive and several negative attributes will always have a dominant attribute: the positive attribute. Similarly, a negative attribute will be the dominant "worst" attribute when it is paired with positive attributes in the profile. When a study contains multiple positive and multiple negative attributes, while no single attribute may be dominant, the "best" will always be chosen from the positive attributes, while the "worst" is chosen from the negative attributes.

We reviewed health related case 2 BWS studies published until 2018 to gain more insights into the type of attributes that have been studied in BWS literature. Details regarding the selection of articles for this scoping review can be found in Appendix A of the electronic supplementary material. Our review identified 87 full-text BWS studies based on a search in PubMed until December 2018 with the search term 'best worst scaling'. For the final data extraction, studies were included when it was an empirical case 2 BWS study, the full-text was available in English, it was health related and it was not a methodological or review study. Eventually, 39 full-texts were included for final data extraction. These 39 studies contained a total of 252 attributes. More than half (n=151, 60%) of the attributes could not be categorized as either positive or negative (in which dominance will not be an issue and models can be estimated). Examples include “start of treatment”, “registration year” and “talk with healthcare provider by phone”. Focusing on the distribution of the 101 positive and negative attributes specifically, most attributes were negative (n=81, 32%), for example “pain”, “malfunction” and “skin injury”. Examples of positive attributes (n=20, 8%) included “simple drug application”, “long duration of efficacy” and “increased life expectancy”. Hence, positive and negative attributes are prevalent in health related BWS studies. From these 39 studies, four contained a mix of positive and negative attributes. All these four studies also included one or more attributes that could not be categorized as positive or negative. Our results suggest that studies that include only a mix of positive and negative attributes will lead to data that suffers from the complete separation problem (Zeng and Zeng, 2019).[vi] This results in estimation failures and as a consequence such studies are unlikely to result in publications.

5.2.2 Choices in case 2 BWS with dominant attributes

To illustrate the problem case 2 BWS encounters with a mix of positive and negative attributes, we introduce a stylized example of a case 2 BWS study containing one positive and three negative attributes. That is, a drug with one benefit and three side effects (i.e., harms). Table 1 shows the four attributes and their levels. Here, A^+ and A^- represent positive and negative attributes. Each attribute has three levels, indicated by subscripts, such that $A^-_{3,2}$ is the second level of the third attribute.

[vi]We want to thank an anonymous reviewer for the insight that the estimation problems stem from an identification problem that results from complete data separation.

Table 1 Case 2 BWS with one positive and three negative attributes

Attributes	Positive or negative	Attribute levels
Attribute 1 (A_1^+) **e.g. efficacy of drug**	+	$A_{1,1}^+$ $A_{1,2}^+$ $A_{1,3}^+$
Attribute 2 (A_2^-) **e.g. side effect 1 of drug**	-	$A_{2,1}^-$ $A_{2,2}^-$ $A_{2,3}^-$
Attribute 3 (A_3^-) **e.g. side effect 2 of drug**	-	$A_{3,1}^-$ $A_{3,2}^-$ $A_{3,3}^-$
Attribute 4 (A_4^-) **e.g. side effect 3 of drug**	-	$A_{4,1}^-$ $A_{4,2}^-$ $A_{4,3}^-$

A single case 2 BWS choice task includes all attributes, each at a specific level, with levels varying across choice tasks. Typical case 2 BWS choice tasks are shown in Figure 2. For each task, individuals will be asked to indicate the "best" and "worst" attribute level based on the attribute and level combinations presented in each choice task. To gain insights into the preferences of individuals, they are repeatedly asked to make "best" and "worst" choices for different choice tasks. If an individual for example selects $A_{1,1}^+$ as "best" and $A_{2,2}^-$ as "worst" in the example in Figure 2, we know that this individual prefers $A_{1,1}^+$ over $A_{2,2}^-$, $A_{3,2}^-$, $A_{4,1}^-$ and also prefers $A_{3,2}^-$ and $A_{4,1}^-$ over $A_{2,2}^-$.

When individuals prefer gains over losses, then, as can be easily observed in Figure 2, every choice task including a mix of positive and negative attributes will contain one or more dominant attributes. In the example in Figure 2 selecting the "best" attribute level requires a choice between a gain and a number of losses, which results in a trivial choice for the dominant attribute when a gain is preferred over losses. Also, the random variation in utility of the dominant attribute will never make it less attractive than the dominated attributes. This differs from the usual situation where an attribute can have a higher utility, on average, but then still the random variation in individuals' utility for the attribute levels typically ensures that other attributes will also sometimes be chosen as "best". More specifically, choices in case 2 BWS with dominant attributes are so simple that respondents do not make mistakes in selecting the dominant attribute level.

Best		Worst
[ ]	$A^{+}_{1,1}$	[ ]
[ ]	$A^{-}_{2,2}$	[ ]
[ ]	$A^{-}_{3,2}$	[ ]
[ ]	$A^{-}_{4,1}$	[ ]

Figure 2 Example choice task case 2 BWS with one positive and three negative attributes

5.3. Model estimation with dominant attributes in case 2 BWS

In this section we review the common modelling approach to case 2 BWS data. We then show how the presence of a dominant attribute leads to infinitely large parameter estimates. Finally, we consider the robustness of this result.

5.3.1 Model based inference for case 2 BWS

Model-based estimation methods for case 2 BWS data are based on utility maximization within the Random Utility Theory (RUT) framework (Louviere et al., 2015). The RUT underpins the models used in a wide array of practical and academic cases to model choice processes (Ben-Akiva and Lerman, 1985; Mcfadden, 2001; McFadden, 1973). In the context of a case 2 BWS choice task, the RUT model is defined as follows: An individual obtains a certain level of utility from each level of each attribute presented in the choice task. For the "best" ("worst") question in case 2 BWS the individual selects the attribute that provides the highest (lowest) utility. There are aspects influencing the utility that the analyst can and cannot capture (Train, 2009). Therefore, the utility for attribute k with level l can be decomposed in two parts. First, a systematic ($V_{k,l}$) part that is common across all choices and respondents, which the analyst can capture. Second, an unobserved residual component ($\varepsilon_{k,l}$), representing the part of utility that cannot be captured by the analyst (unobserved utility component). In this paper we do not restrict the systematic part of utility and use, without loss of generality, the notation $V_{k,l} = f_k(A_{k,l})$, as our focus is not on the specific functional representations of the utility functions and with f_k representing different attribute specific functional utility forms.

Considering the analyst only observes choices and not the underlying true utility levels, probabilistic models are used to account for the unobserved utility component when analyzing choice data (Train, 2009). This results in a probability, in the situation of two attributes for example, of selecting attribute k with level l over attribute m with level n given by:

$$P(best = A_{k,l}) = P\left(V_{k,l} + \varepsilon_{k,l} > V_{m,n} + \varepsilon_{m,n}\right) \quad \text{[eq.1]}$$

The multinomial logit (MNL) model and its generalizations are the common probabilistic model to analyze case 2 BWS choice data (Hawkins et al., 2019; Mühlbacher et al., 2016a). MNL estimation of case 2 BWS data can be performed in two ways, depending on the assumed psychological processes of decision-making by which individuals decide about their "best" and "worst" choices (Louviere et al., 2015): First, the maximum difference model (maxdiff), in which individuals choose that best-worst pair that maximizes the utility difference between “best” and “worst”. Second, in the sequential model individuals make their “best” and “worst” choices in two stages: first choosing the “best” (“worst”) from all options and then choosing the “worst” (“best”) from all remaining options. In this section we focus on the “best”-first sequential model to elaborate on the estimation problems of case 2 BWS data with dominant attributes, though the same issues arise with the maxdiff approach or if the “worst” option is selected first followed by selecting the best option (see section 3.3).

We follow the common assumption underlying the MNL model, which is that the error term is independently and identically distributed (IID) and extreme value (EV) type I across alternatives (McFadden, 1973; Mühlbacher et al., 2016a; Train, 2009). This results in the probability that within a specific choice task *s* an individual selects attribute k with level $l_{k,s}$ as “best”, given *M* attributes (*m*) with choice task *s* specific attribute levels $n_{m,s}$, given by (Flynn and Marley, 2014):

$$P(best = A_{k,l_{k,s}}) = \frac{exp\left(V_{k,l_{k,s}}\right)}{\sum_{m=1}^{M} exp\left(V_{m,n_{m,s}}\right)} \quad \text{[eq.2]}$$

5.3.2 Dominant attributes in case 2 BWS

Returning to our case 2 BWS example with one positive and three negative attributes in Table 1, where individuals always select the dominant, positive attribute as “best”. In terms of the utility of the positive and negative attribute levels, this implies that the utility of the positive attribute is always greater than the utility of the negative attributes, hence we know that for all l, m and n:

$$V(A^{+}_{k,l}) + \varepsilon^{+}_{k,l} > V(A^{-}_{m,n}) + \varepsilon^{-}_{m,n} \quad \text{[eq.3]}$$

with $V(A_{k,l}^{+})$ and $V(A_{m,n}^{-})$ representing the systematic utilities for the positive and negative attributes and $\varepsilon_{k,l}^{+}$ and $\varepsilon_{m,n}^{-}$ representing the associated unobserved utility components. In this situation with a dominant positive attribute, the inequality in equation 3 needs to hold for all possible values of $\varepsilon_{k,l}^{+}$ and $\varepsilon_{m,n}^{-}$. Given the unbounded support of the extreme value distribution, this can only be the case when the difference in utilities between the positive and negative attributes, $V(A_{k,l}^{+}) - V(A_{m,n}^{-})$ is infinite.

To demonstrate the estimation problem in another way, let us focus on the case 2 BWS example in Figure 2. Individuals will always select $A_{1,1}^{+}$ as best. Within the MNL model specification, the probability that an individual selects the positive attribute as "best" is given by:

$$P(best = A_{1,1}^{+}) = \frac{exp\left(V(A_{1,1}^{+})\right)}{exp\left(V(A_{1,1}^{+})\right) + exp\left(V(A_{2,1}^{-})\right) + exp\left(V(A_{3,2}^{-})\right) + exp\left(V(A_{4,1}^{-})\right)} \quad \text{[eq.4]}$$

Since the positive attribute will always be selected as "best", this implies that $P(best = A_{1,1}^{+}) = 1$.

As shown before, mixing positive and negative attributes will likely lead to attribute level dominance and therefore to data that suffers from the complete separation problem: meaning that the positive attribute is always selected as "best". This will lead to corresponding estimation problems when using the MNL model for estimation. In other words: the MNL model is inconsistent with the assumption of dominance of an attribute level in BWS-2 as that results in the complete separation problem and therefore to estimation problems when fitting the MNL model (Zeng and Zeng, 2019). The aim of this study is to show that a mix of positive and negative attributes in BWS-2 will lead to attribute level dominance, which results in identification problems that leads to failure of the MNL model.

In the simulation part of this study, complete separation is induced by imposing individuals to always select a positive above a negative attribute when selecting the “best” attribute.

In case 2 BWS all attributes are assumed to be measured on the same scale and modelling case 2 BWS data always requires the analyst to select one attribute level to be set as the reference level (Potoglou et al., 2011; van Dijk et al., 2016). Without loss of generality, we use level 1 of attribute 2 ($A^-_{2,1}$) as the reference level, with $V(A^-_{2,1})=0$ leading to $exp(V(A^-_{2,1}))=1$. Since $exp(V(A^+_{1,1}))$ is both in the nominator and denominator of the MNL model specification in equation 4, $exp(V(A^-_{2,1}))$ contributes 1 to the denominator and both $exp(V(A^-_{3,2}))$ and $exp(V(A^-_{4,1}))$ contribute non-negative amounts, the probability will be smaller than one. Therefore, the only way the MNL probability of selecting $A^+_{1,1}$ as “best” will be equal to one requires $V(A^+_{1,1})$ to become infinitely large. That way the utility values of the other attribute levels have essentially no impact. An infinitely large $V(A^+_{1,1})$ prevents the estimation procedure from converging, effectively leading to a situation where we will not be able to estimate the MNL parameters.

5.3.3 Robustness of argumentation

In the examples above, we focused on a scenario with one positive and three negative attributes. The same problem manifests when there are two or more positive attributes in combination with one or more negative attributes. The difference in utilities between any of the positive attributes and the set of negative attributes must be infinitely large to ensure an individual selects one of the positive attributes as “best” (equation 3), resulting in the same type of estimation problems. In general, a mix of any number of positive and negative attributes leads to estimation problems. In section 4.2 we will also show this using simulated data.

A similar situation obtains with the maxdiff model. Focusing on pairs of attribute level combinations rather than individual attribute levels relative to the other attribute levels, based on the utility functions described in equation 3, also requires infinitely large differences between the utility levels of the positive and negative attributes, in this case within pairs of such attributes.

Finally, we consider other statistical models. The mixed logit model (MXL) is often used in choice modelling to accommodate for heterogeneity of preferences (Train, 2009). The latent class model (LCM) also accommodates for heterogeneity of preferences by sorting individuals in classes (Train, 2009). The arguments above apply at both the individual and population levels, so accounting for heterogeneity of preferences or aggregating to the population level with these models will not alleviate the estimation problem.

5.4 Simulation study

Section 5.3 analytically showed how the presence of a dominant attribute in case 2 BWS leads to infinitely large parameter values. In this section we will show this estimation problem making use of simulated data.

5.4.1 Simulation design

We simulated the example from Table 1: one positive (A^+) and three negative (A^-) attributes, each with three attribute levels. To clearly show the impact of dominance on parameter estimates, we focused on two different simulation scenarios. In the first scenario, the positive attribute is generally preferred over the negative attributes, but it is not dominant, i.e. it is not always selected as “best” (no-dominance scenario); hence, $P(best = A^+) < 1$ in equation 4. In the second scenario, the positive attribute is dominant, so it is not only preferred on average, but it is always selected as “best” (dominance scenario); i.e., $P(best = A^{+}) = 1$. The precise utility values for each attribute level are presented in Table 2.

Both scenarios used the same orthogonal main effect plan (OMEP) experimental design with 9 choice tasks from Hahn & Shapiro (1966). Based on the number of attributes and levels, the OMEP catalogue provided us with the information that in order to get an orthogonal design 9 choice tasks were needed. The specific combination of attribute levels in each choice task could also be found in this catalogue. Both scenarios also used MNL for model estimation, using a maxdiff approach. Since the aim of our study was to investigate the effect of mixing positive and negative attributes – leading to dominance – on BWS case 2 outcomes, we adjusted the data generating process (DGP) such that the positive attribute was always selected as “best” (or one of the two positive attributes in the case with multiple positive attributes). A population sample size of 1000 with results accumulated over 500 simulated replications was used (Koehler et al., 2009). The simulation code was written in Julia programming language version 1.0.3 (https://julialang.org/).

Table 2 Utility values for each attribute level in the two scenarios

Attributes	Attribute levels			Utility levels		
Attribute 1 (A_1^+)	$A_{1,1}^+$	$A_{1,2}^+$	$A_{1,3}^+$	1.00	1.50	2.00
Attribute 2 (A_2^-)	$A_{2,1}^-$	$A_{2,2}^-$	$A_{2,3}^-$	0.00	-0.50	-1.00
Attribute 3 (A_3^-)	$A_{3,1}^-$	$A_{3,2}^-$	$A_{3,3}^-$	-1.00	-1.00	-1.00
Attribute 4 (A_4^-)	$A_{4,1}^-$	$A_{4,2}^-$	$A_{4,3}^-$	-1.00	-1.50	-2.00

5.4.2 Simulation results

Table 3 shows the average estimated utility values for two scenarios, differing in whether the positive attribute was dominant or not, across the 500 simulation runs. In the no-dominance scenario, we were able to estimate the true utilities for both the positive and the negative attributes (with ($A^{-}_{2,1}$ set as reference level). However, in the dominance scenario, where the positive attribute was always preferred over the negative attributes, the estimated values for the positive attribute levels are very large and unrelated to the true utilities, as predicted by our analytical derivations. The utility levels for the negative attributes, were properly recovered, but the estimates are somewhat less precise in the dominance scenario, relative to the no-dominance scenario.

Table 3 Mean estimated utility and SD values for each attribute level from 500 simulated replications for the two scenarios

	No-dominance		Dominance		True values
Attribute level	Estimated value	(SD)	Estimated value	(SD)	
$A^{+}_{1,1}$ (beta1)	1.00	(0.05)	37.23	(3.37)	**1.00**
$A^{+}_{1,2}$ (beta2)	1.50	(0.05)	38.87	(3.40)	**1.50**
$A^{+}_{1,3}$ (beta3)	2.00	(0.05)	37.56	(3.36)	**2.00**
$A^{-}_{2,1}$ (Ref)	0.00	Ref	0.00	Ref	**0.00**
$A^{-}_{2,2}$ (beta4)	-0.50	(0.05)	-0.53	(0.08)	**-0.50**
$A^{-}_{2,3}$ (beta5)	-1.00	(0.05)	-0.99	(0.07)	**-1.00**
$A^{-}_{3,1}$ (beta6)	-1.00	(0.05)	-1.00	(0.07)	**-1.00**
$A^{-}_{3,2}$ (beta7)	-1.00	(0.05)	-1.00	(0.07)	**-1.00**
$A^{-}_{3,3}$ (beta8)	-1.00	(0.05)	-1.01	(0.07)	**-1.00**
$A^{-}_{4,1}$ (beta9)	-1.00	(0.05)	-0.99	(0.06)	**-1.00**
$A^{-}_{4,2}$ (beta10)	-1.50	(0.05)	-1.49	(0.07)	**-1.50**
$A^{-}_{4,3}$ (beta11)	-2.00	(0.06)	-2.00	(0.09)	**-2.00**

The histograms in Figure 3 show the distribution of the estimated utility levels for attribute 1 in the two scenarios. The vertical white dashed lines indicate the true utility values from Table 2. Overall, we were able to infer the true utilities back for the positive attribute in the no-dominance scenario (plots a-c). However, in the scenario with dominance (plots d-f), the histograms show that the estimated utilities for the positive attributes become very large and hence the true utility values were not recovered. Not only are these large parameter estimates non-informative because of their large value, but the dispersion of these estimates is also much larger.

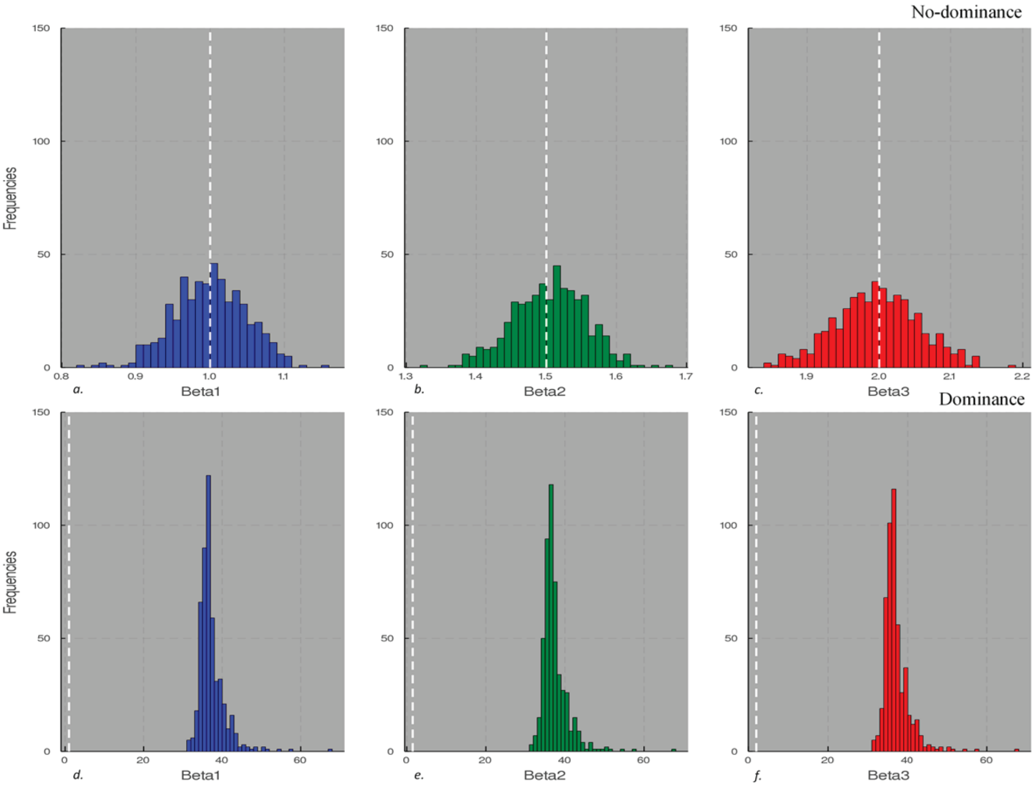


Figure 3 Distribution of estimated utility values for attribute 1 V(A_1^+) and its levels in the simulations no-dominance (a-c) and dominance (d-f). Dashed white lines indicate the true utility values.

Dominance even affects the relative ranking of the positive attribute levels. Based on the utility values from Table 2 the correct ranking for the positive attribute levels is $A^{+}_{1,3}$ > $A^{+}_{1,2}$ > $A^{+}_{1,1}$. Table 4 presents the number of times the attribute levels had the correct rank based on the parameter estimates. The inferred ranking is fully aligned with the underlying data generating process in the no-dominance scenario, while it clearly fails to reflect the correct ranking in the dominance scenario. A single positive dominant attribute makes it impossible to infer the preference order of the levels of this attribute.

Table 4 Number of times (percentages) attribute levels are ranked properly based on point estimates of the parameters for the positive attribute levels in scenarios no-dominance and dominance

Attribute levels (true rank)	**No-dominance**	**Dominance**
$A^{+}_{1,1}$ **(rank 3)**	500 (100%)	9 (2%)
$A^{+}_{1,2}$ **(rank 2)**	500 (100%)	9 (2%)
$A^{+}_{1,3}$ **(rank 1)**	500 (100%)	418 (84%)

To study what happens when there are multiple positive and negative attributes, we designed a second simulation study. In this simulation, we added a second positive attribute to the setting of the previous simulation study. The utility values for the positive attribute that was added are set at: 0.5 ($A^{+}_{2,1}$), 0.625 ($A^{+}_{2,2}$) and 0.75 ($A^{+}_{2,3}$). Figure 4 shows the distribution of the estimated utility values for attributes 1 and 2 for these scenarios. In the no-dominance scenario, the true utility values were recovered well when looking at the mean estimated utility values. As expected, with dominance the utility estimates for the positive attributes are very large and distant from the true utility values, demonstrating that also dominance causes estimation problems with a mix of multiple positive and negative attributes. Unlike the case with a single positive attribute, the relative rankings for the positive attributes is correctly retrieved. This is because in this simulation design there are two positive attributes that can be selected as “best” instead of one, providing comparisons consistent with the model assumptions, which enables recovery of the actual rankings. This means that depending on the study goal, that is to determine rankings or to compute willingness-to-pay (WTP) for example, dominance is expected to impact outcomes. This is especially the case when dealing with trade-offs between gains and losses in for example WTP computations.

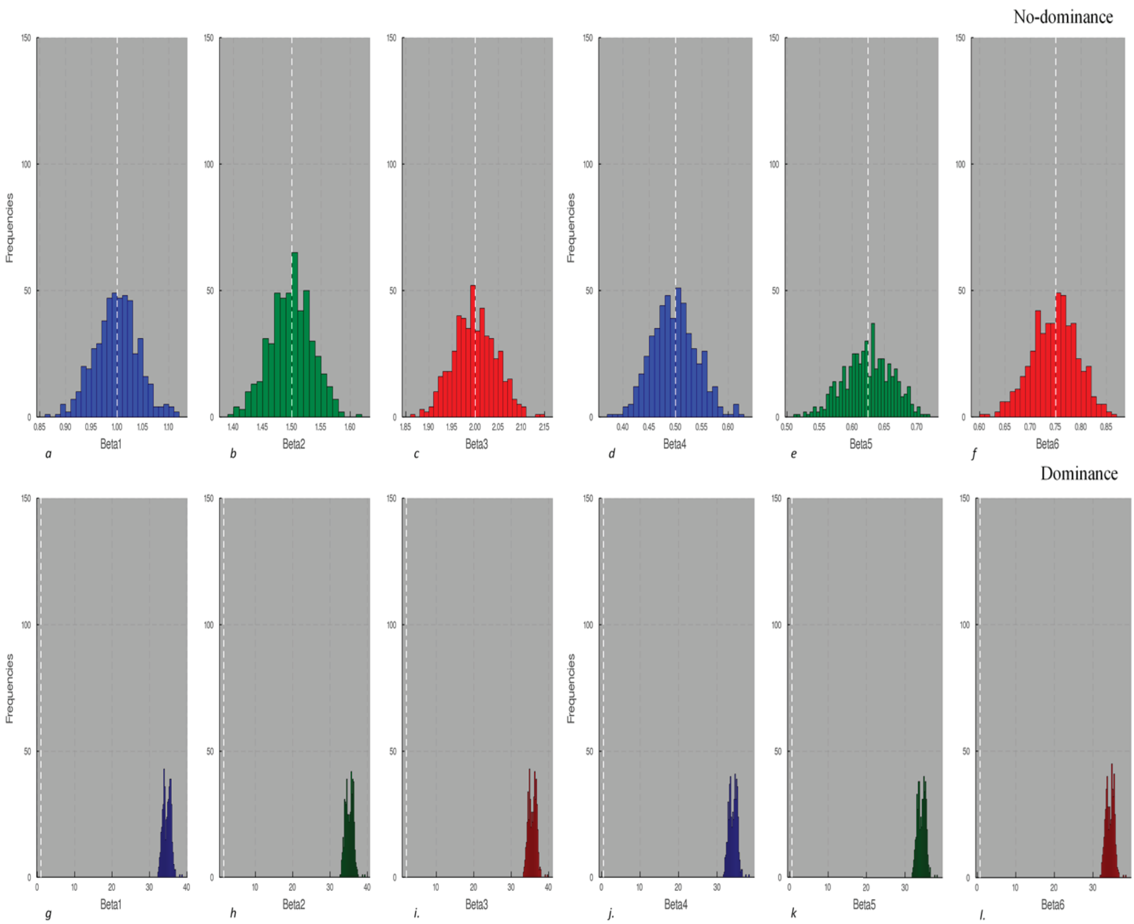


Figure 4 Distribution of estimated utility values for the two positive attributes 1 V(A_1^+) and 2 V(A_2^+) in the no-dominance (a-f) and dominance scenarios (g-l). Dashed white lines indicate the true utility values.

5.5 Discussion

In this paper we studied how using a mixture of positive and negative attributes affects the performance of a logit-model-based analysis of case 2 BWS data. Our analysis relies on a single assumption on respondents' preferences, which is that individuals will always prefer and select a positive attribute above a negative attribute when selecting the "best" attribute.[vii] This assumption is grounded in behavioral economics (Levin and Gaeth, 1988; Tversky and Kahneman, 1981), but also aligns well with common sense. People will prefer health improvements over side effects like headaches or nausea, when given the choice. We showed analytically that BWS experiments containing a mix of both positive and negative attributes results in infinitely large differences in utilities between the positive and negative attributes due to the complete separation problem in the data. Based on these findings we predict that model estimation will fail when using data from a case 2 BWS study with a mix of positive and negative attributes.

Simulation results confirmed the analytical predictions. In particular, the difference between the utility estimates for the positive and the negative attribute(s) was much larger than the corresponding difference between the true values. When there was only a single positive attribute, we were not even able to recover the relative preference ordering for its levels in our simulations, and the same will occur when there is a single negative attribute and multiple positive attributes. Once multiple positive and multiple negative attributes are combined, the relative ordering of the attributes within the set of positive attributes and within the set of negative attributes is correctly assessed. These two parts of the utility scale, however, are at a large distance and not necessarily on a comparable scale, e.g. when choosing the "best" among the positive attributes is more difficult than choosing the "worst"

[vii] As mentioned in section 3.2, MNL is inconsistent with attribute level dominance as it leads to completely separated data and the corresponding identification problems. To show this, we imposed preference of the positive attribute over the negative attributes, which induces complete data separation. to illustrate the fundamental identification and estimation problems that arise from BWS-2 studies that include a mix of positive and negative attributes.

among the negative attributes. This is an important finding, as it raises concerns in studies that contain both positive and negative attributes, e.g. when computing willingness to pay with a negative attribute, e.g. costs, for a positive attribute, e.g. a health benefit. To avoid issues of dominance in case 2 BWS experiments, we can frame all attribute levels to have the same "sign", either all positive or negative. A "degree of recovery" can be translated into "degree of condition remaining" or "probability of side effects occurring" can be reframed as "probability of side effects being absent". By reframing the attributes all positive or negative (e.g., chance of not being cured in Figure 1b), we can avoid the dominance-related issues of case 2 BWS. However, it is an open question whether people interpret these two attribute frames in the same manner.

Although, our study is the first step in understanding case 2 BWS issues regarding the type of attributes to include in such choice experiments, there are still several open questions. First, only a few studies that were identified with the literature review included a mixture of positive and negative attributes. This raises the question how relevant the issue of attribute dominance is. It could however be the case that studies that result in estimation problems are unlikely to result in publications. The fact that there are not many studies reporting these issues does not mean this problem should be ignored. This study therefore tries to inform and warn case 2 BWS users when designing choice tasks, since similar questions have been raised when using DCEs in specific situations (Flynn et al., 2008). Second, our definition of positive and negative attributes as gains and losses respectively deserves empirical scrutiny. Future research might focus on reference points to which attribute levels are compared to as a driver of attribute level preferences. Third, in this paper we analyzed dominance at the level of an attribute (all levels of the positive attribute are dominant). However, one can imagine that dominance in case 2 BWS also occurs for a specific attribute level, e.g. the highest efficacy level is always preferred over all other attribute levels, or the highest price level is always considered worse than all other attribute levels.

To conclude, case 2 BWS with a mix of positive and negative attributes leads to dominance related problems in model estimation. Nonetheless, we believe that case 2 BWS holds the potential of being valuable for eliciting preferences, if only good (positive) or only bad (negative) attributes are included in the experiment, but not for both.

References

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis : theory and application to travel demand. MIT Press.

Bliemer, M.C.J., Rose, J.M., 2011. Experimental design influences on stated choice outputs: An empirical study in air travel choice. Transp. Res. Part A Policy Pract. 45, 63–79. https://doi.org/10.1016/j.tra.2010.09.003

Bliemer, M.C.J., Rose, J.M., Chorus, C.G., 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. Transp. Res. Part B Methodol. 102, 83–104. https://doi.org/10.1016/j.trb.2017.05.005

Flynn, T.N., 2010. Valuing citizen and patient preferences in health: Recent developments in three types of best-worst scaling. Expert Rev. Pharmacoeconomics Outcomes Res. 10, 259–267. https://doi.org/10.1586/erp.10.29

Flynn, T.N., Louviere, J.J., Marley, A.A.J., Coast, J., Peters, T.J., 2008. Rescaling quality of life values from discrete choice experiments for use as QALYs: A cautionary tale. Popul. Health Metr. 6, 1–11. https://doi.org/10.1186/1478-7954-6-6

Flynn, T.N., Louviere, J.J., Peters, T.J., Coast, J., 2007. Best-worst scaling: What it can do for health care research and how to do it. J. Health Econ. 26, 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002

Flynn, T.N., Marley, A.A.J., 2014. Best-worst scaling: theory and methods. Handb. Choice Model. 178–201.

Hahn, G.J., Shapiro, S.S., 1966. A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. General Electric Research and Development Center, Schenectady New York.

Hawkins, G.E., Islam, T., Marley, A.A.J., 2019. Like it or not, you are using one value representation. Decision 6, 237–260. https://doi.org/10.1037/dec0000100

Hollin, I.L., Peay, H.L., Apkon, S.D., Bridges, J.F.P., 2017. Patient-centered benefit-risk assessment in duchenne muscular dystrophy. Muscle Nerve 55, 626–634. https://doi.org/10.1002/mus.25411

Huber, J., Payne, J.W., Puto, C., 1982. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. J. Consum. Res. 9, 90. https://doi.org/10.1086/208899

Kahneman, D., Tversky, A., 1988. Prospect theory: An analysis of decision under risk, in: Decision, Probability and Utility. Cambridge University Press, pp. 183–214. https://doi.org/10.1017/CBO9780511609220.014

Koehler, E., Brown, E., Haneuse, S.J.P.A., 2009. On the assessment of Monte Carlo error in simulation-based Statistical analyses. Am. Stat. 63, 155–162. https://doi.org/10.1198/tast.2009.0030

Krucien, N., Watson, V., Ryan, M., 2017. Is Best–Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments. Heal. Econ. (United Kingdom) 26, e1–e16. https://doi.org/10.1002/hec.3459

Levin, I.P., Gaeth, G.J., 1988. How Consumers are Affected by the Framing of Attribute Information Before and After Consuming the Product. J. Consum. Res. 15, 374. https://doi.org/10.1086/209174

Louviere, J.J., Flynn, T.N., Marley, A.A.J., 2015. Best-worst scaling: Theory, methods and applications, Best-Worst Scaling: Theory, Methods and Applications. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107337855

Mcfadden, D., 2001. Economic Choices, The American Economic Review.

McFadden, D., 1973. Conditional Logit Analysis of Qualitative Choice Behaviour, Frontiers in Econometrics. Academic Press, New York. https://doi.org/10.1108/eb028592

Mühlbacher, A.C., Kaczynski, A., Zweifel, P., Johnson, F.R., 2016a. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ. Rev. 6, 1–14. https://doi.org/10.1186/s13561-015-0079-x

Mühlbacher, A.C., Zweifel, P., Kaczynski, A., Johnson, F.R., 2016b. Experimental measurement of preferences in health care using best-worst scaling (BWS): theoretical and statistical issues. Health Econ. Rev. 6, 5. https://doi.org/10.1186/s13561-015-0077-z

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

Severin, F., Schmidtke, J., Mühlbacher, A., Rogowski, W.H., 2013. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. Eur. J. Hum. Genet. 21, 1202–1208. https://doi.org/10.1038/ejhg.2013.36

Tarini, B.A., Simon, N.J., Payne, K., Gebremariam, A., Rose, A., Prosser, L.A., 2018. An Assessment of Public Preferences for Newborn Screening Using Best–Worst Scaling. J. Pediatr. 201, 62-68.e1. https://doi.org/10.1016/j.jpeds.2018.05.035

Tervonen, T., Schmidt-Ott, T., Marsh, K., Bridges, J.F.P., Quaife, M., Janssen, E., 2018. Assessing Rationality in Discrete Choice Experiments in Health: An Investigation into the Use of Dominance Tests. Value Heal. 21, 1192–1197. https://doi.org/10.1016/j.jval.2018.04.1822

Thurstone, L.L., 1927. A law of comparative judgment. Psychol. Rev. 34, 273–286. https://doi.org/10.1037/h0070288

Train, K.E., 2009. Discrete choice methods with simulation, second edition, Discrete Choice Methods with Simulation, Second Edition. Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Tversky, A., Kahneman, D., 1981. The Framing of Decisions and the Psychology of Choice.

van Dijk, J.D., Groothuis-Oudshoorn, C.G.M., Marshall, D.A., IJzerman, M.J., 2016. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. Value Heal. 19, 316–322. https://doi.org/10.1016/j.jval.2015.12.020

Whitty, J.A., Ratcliffe, J., Chen, G., Scuffham, P.A., 2014. Australian public preferences for the funding of new health technologies: A comparison of discrete choice and profile case best-worst scaling methods. Med. Decis. Mak. 34, 638–654. https://doi.org/10.1177/0272989X14526640

Zeng, G., Zeng, E., 2019. On the relationship between multicollinearity and separation in logistic regression. Commun. Stat. Simul. Comput. https://doi.org/10.1080/03610918.2019.1589511

Appendix A

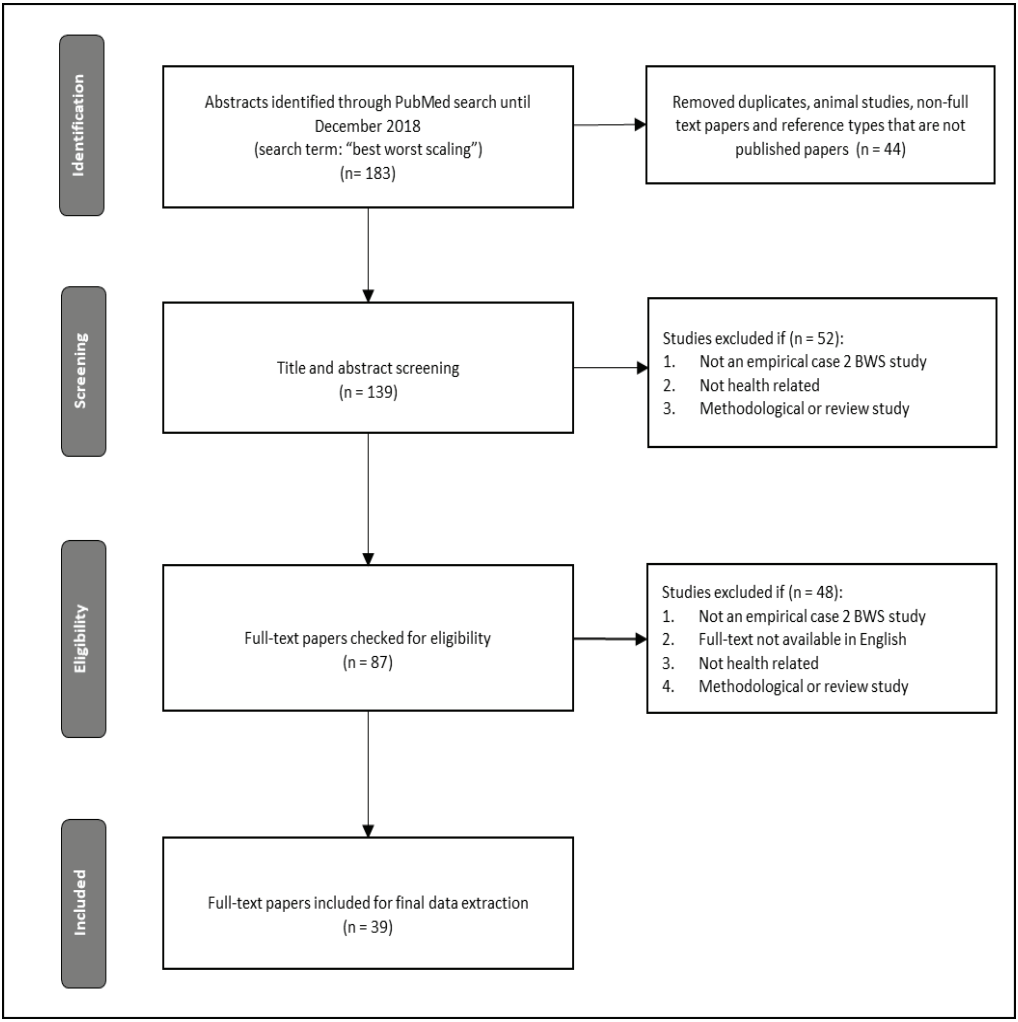


Figure A Flow diagram of scoping review to identify case 2 BWS studies

Full-text papers included for data extraction

1. Kok, M. M., Weernink, M. G. M., von Birgelen, C., Fens, A., van der Heijden, L. C., & van Til, J. A. (2018). Patient preference for radial versus femoral vascular access for elective coronary procedures: The PREVAS study. Catheterization and Cardiovascular Interventions : Official Journal of the Society for Cardiac Angiography & Interventions, 91(1), 17–24. https://doi.org/10.1002/ccd.27039
2. Richardson, J. S., Messonnier, M. L., & Prosser, L. A. (2018). Preferences for health economics presentations among vaccine policymakers and researchers. Vaccine, 36(43), 6416–6423. https://doi.org/10.1016/j.vaccine.2018.08.049
3. Tarini, B. A., Simon, N. J., Payne, K., Gebremariam, A., Rose, A., & Prosser, L. A. (2018). An Assessment of Public Preferences for Newborn Screening Using Best–Worst Scaling. Journal of Pediatrics, 201, 62-68.e1. https://doi.org/10.1016/j.jpeds.2018.05.035
4. Castillo, W. C., Ross, M., Tariq, S., & Dos Reis, S. (2018). Best-worst scaling to prioritize outcomes meaningful to caregivers of youth with mental health multimorbidities: A pilot study. Journal of Developmental and Behavioral Pediatrics, 39(2), 101–108. https://doi.org/10.1097/DBP.0000000000000525
5. Ross, M., Nguyen, V., Bridges, J. F. P., Ng, X., Reeves, G., & Frosch, E. (2018). Caregivers ' Priorities and Observed Outcomes of Attention-. J Dev Behav Pediatr., 39(2), 93–100.
6. Ng, X., Bridges, J. F. P., Ross, M. M., Frosch, E., Reeves, G., Cunningham, C. E., & dosReis, S. (2017). A Latent Class Analysis to Identify Variation in Caregivers' Preferences for their Child's Attention-Deficit/Hyperactivity Disorder Treatment: Do Stated Preferences Match Current Treatment? Patient, 10(2), 251–262. https://doi.org/10.1007/s40271-016-0202-z
7. dosReis, S., Park, A., Ng, X., Frosch, E., Reeves, G., Cunningham, C., … Bridges, J. F. P. (2017). Caregiver Treatment Preferences for Children with a New Versus Existing Attention-Deficit/Hyperactivity Disorder Diagnosis. Journal of Child and Adolescent Psychopharmacology, 27(3), 234–242. https://doi.org/10.1089/cap.2016.0157
8. Khakban, A., Mohammadi, T., Lynd, L. D., Mabbott, D., Bouffet, E., Gastonguay, L., … Marra, C. A. (2017). Societal preferences in the treatment of pediatric medulloblastoma: Balancing risk of death and quality of life. Pediatric Blood & Cancer, 64(6). https://doi.org/10.1002/pbc.26340
9. Meregaglia, M., Cairns, J., Alfieri, S., Favales, F., Mazzitelli, D., Orlandi, E., … Bossi, P. (2017).

Eliciting Preferences for Clinical Follow-Up in Patients with Head and Neck Cancer Using Best-Worst Scaling. Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research, 20(6), 799–808. https://doi.org/10.1016/j.jval.2017.01.012

10. Hollin, I. L., Peay, H. L., Apkon, S. D., & Bridges, J. F. P. (2017). Patient-centered benefit-risk assessment in duchenne muscular dystrophy. Muscle & Nerve, 55(5), 626–634. https://doi.org/10.1002/mus.25411
11. Huynh, E., Coast, J., Rose, J., Kinghorn, P., & Flynn, T. (2017). Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. Social Science & Medicine (1982), 189, 114–128. https://doi.org/10.1016/j.socscimed.2017.07.012
12. Beusterien, K., Chan, E., Such, P., de Jong Laird, A., Heres, S., Amos, K., … Bridges, J. F. P. (2017). Development of a stated-preference instrument to prioritize treatment goals in recent onset schizophrenia. Current Medical Research and Opinion, 33(12), 2129–2136. https://doi.org/10.1080/03007995.2017.1384717
13. Weernink, M. G. M., van Til, J. A., Groothuis-Oudshoorn, C. G. M., & IJzerman, M. J. (2017). Patient and Public Preferences for Treatment Attributes in Parkinson's Disease. The Patient, 10(6), 763–772. https://doi.org/10.1007/s40271-017-0247-7
14. Howell, M., Wong, G., Rose, J., Tong, A., Craig, J. C., & Howard, K. (2017). Patient Preferences for Outcomes After Kidney Transplantation: A Best-Worst Scaling Survey. Transplantation, 101(11), 2765–2773. https://doi.org/10.1097/TP.0000000000001793
15. Ratcliffe, J., Huynh, E., Stevens, K., Brazier, J., Sawyer, M., & Flynn, T. (2016). Nothing about Us Without Us? A Comparison of Adolescent and Adult Health-State Values for the Child Health Utility-9D Using Profile Case Best-Worst Scaling. Health Economics (United Kingdom), 25(4), 486–496. https://doi.org/10.1002/hec.3165
16. Howell, M., Wong, G., Rose, J., Tong, A., Craig, J. C., & Howard, K. (2016). Eliciting patient preferences, priorities and trade-offs for outcomes following kidney transplantation: a pilot best-worst scaling survey. BMJ Open, 6(1), e008163. https://doi.org/10.1136/bmjopen-2015-008163
17. Pooripussarakul, S., Riewpaiboon, A., Bishai, D., Muangchana, C., & Tantivess, S. (2016). What criteria do decision makers in Thailand use to set priorities for vaccine introduction? BMC Public Health, 16, 684. https://doi.org/10.1186/s12889-016-3382-5
18. Pollitt, A., Potoglou, D., Patil, S., Burge, P., Guthrie, S., King, S., … Grant, J. (2016). Understanding

the relative valuation of research impact: a best-worst scaling experiment of the general public and biomedical and health researchers. BMJ Open, 6(8), e010916. https://doi.org/10.1136/bmjopen-2015-010916

19. Ratcliffe, J., Huynh, E., Chen, G., Stevens, K., Swait, J., Brazier, J., … Flynn, T. (2016). Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Social Science and Medicine, 157, 48–59. https://doi.org/10.1016/j.socscimed.2016.03.042
20. Coast, J., Huynh, E., Kinghorn, P., & Flynn, T. (2016). Complex Valuation: Applying Ideas from the Complex Intervention Framework to Valuation of a New Measure for End-of-Life Care. PharmacoEconomics, 34(5), 499–508. https://doi.org/10.1007/s40273-015-0365-9
21. Weernink, M. G. M., van Til, J. A., van Vugt, J. P. P., Movig, K. L. L., Groothuis-Oudshoorn, C. G. M., & IJzerman, M. J. (2016). Involving Patients in Weighting Benefits and Harms of Treatment in Parkinson’s Disease. PloS One, 11(8), e0160771. https://doi.org/10.1371/journal.pone.0160771
22. Kiltz, U., Essers, I., Hiligsmann, M., Braun, J., Maksymowych, W. P., Taylor, W. J., … Boonen, A. (2016). Which aspects of health are most important for patients with spondyloarthritis? A Best Worst Scaling based on the ASAS Health Index. Rheumatology (United Kingdom), 55(10), 1771–1776. https://doi.org/10.1093/rheumatology/kew238
23. dosReis, S., Ng, X., Frosch, E., Reeves, G., Cunningham, C., & Bridges, J. F. P. (2015). Using Best-Worst Scaling to Measure Caregiver Preferences for Managing their Child’s ADHD: A Pilot Study. The Patient, 8(5), 423–431. https://doi.org/10.1007/s40271-014-0098-4
24. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. Health Economics (United Kingdom), 24(3), 258–269. https://doi.org/10.1002/hec.3014
25. Ungar, W. J., Hadioonzadeh, A., Najafzadeh, M., Tsao, N. W., Dell, S., & Lynd, L. D. (2015). Parents and adolescents preferences for asthma control: a best-worst scaling choice experiment using an orthogonal main effects design. BMC Pulmonary Medicine, 15, 146. https://doi.org/10.1186/s12890-015-0141-9
26. O’Hara, N. N., Roy, L., O’Hara, L. M., Spiegel, J. M., Lynd, L. D., FitzGerald, J. M., … Marra, C. A. (2015). Healthcare Worker Preferences for Active Tuberculosis Case Finding Programs in South Africa: A Best-Worst Scaling Choice Experiment. PloS One, 10(7), e0133304.

https://doi.org/10.1371/journal.pone.0133304

27. Franco, M. R., Howard, K., Sherrington, C., Ferreira, P. H., Rose, J., Gomes, J. L., & Ferreira, M. L. (2015). Eliciting older people's preferences for exercise programs: a best-worst scaling choice experiment. Journal of Physiotherapy, 61(1), 34–41. https://doi.org/10.1016/j.jphys.2014.11.001
28. Ungar, W. J., Hadioonzadeh, A., Najafzadeh, M., Tsao, N. W., Dell, S., & Lynd, L. D. (2014). Quantifying preferences for asthma control in parents and adolescents using best-worst scaling. Respiratory Medicine, 108(6), 842–851. https://doi.org/10.1016/j.rmed.2014.03.014
29. Xu, F., Chen, G., Stevens, K., Zhou, H., Qi, S., Wang, Z., … Ratcliffe, J. (2014). Measuring and valuing health-related quality of life among children and adolescents in mainland China--a pilot study. PloS One, 9(2), e89222. https://doi.org/10.1371/journal.pone.0089222
30. Damery, S., Biswas, M., Billingham, L., Barton, P., Al-Janabi, H., & Grimer, R. (2014). Patient preferences for clinical follow-up after primary treatment for soft tissue sarcoma: a cross-sectional survey and discrete choice experiment. European Journal of Surgical Oncology : The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 40(12), 1655–1661. https://doi.org/10.1016/j.ejso.2014.04.020
31. Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.-M., Brazier, J., … Wall, B. (2012). Outcomes of social care for adults: developing a preference-weighted measure. Health Technology Assessment (Winchester, England), 16(16), 1–166. https://doi.org/10.3310/hta16160
32. Molassiotis, A., Emsley, R., Ashcroft, D., Caress, A., Ellis, J., Wagland, R., … Blackhall, F. (2012). Applying Best-Worst scaling methodology to establish delivery preferences of a symptom supportive care intervention in patients with lung cancer. Lung Cancer (Amsterdam, Netherlands), 77(1), 199–204. https://doi.org/10.1016/j.lungcan.2012.02.001
33. Ratcliffe, J., Flynn, T., Terlich, F., Stevens, K., Brazier, J., & Sawyer, M. (2012). Developing Adolescent-Specific Health State Values for Economic Evaluation. PharmacoEconomics, 30(8), 713–727. https://doi.org/10.2165/11597900-000000000-00000
34. Al-Janabi, H., Flynn, T. N., & Coast, J. (2011). Estimation of a preference-based carer experience scale. Medical Decision Making : An International Journal of the Society for Medical Decision Making, 31(3), 458–468. https://doi.org/10.1177/0272989X10381280
35. Günther, O. H., Kürstein, B., Riedel-Heller, S. G., & König, H. H. (2010). The role of monetary and nonmonetary incentives on the choice of practice establishment: A stated preference study of young

physicians in Germany. Health Services Research, 45(1), 212–229. https://doi.org/10.1111/j.1475-6773.2009.01045.x

36. Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J., & Peters, T. J. (2008). Valuing the ICECAP capability index for older people. Social Science and Medicine, 67(5), 874–882. https://doi.org/10.1016/j.socscimed.2008.05.015

37. Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. Journal of Health Economics, 26(1), 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002

38. Coast, J., Flynn, T. N., Salisbury, C., Louviere, J., & Peters, T. J. (2006). Maximising responses to discrete choice experiments: A randomised trial. Applied Health Economics and Health Policy, 5(4), 249–260. https://doi.org/10.2165/00148365-200605040-00006

39. Coast, J., Salisbury, C., De Berker, D., Noble, A., Horrocks, S., Peters, T. J., & Flynn, T. N. (2006). Preferences for aspects of a dermatology consultation. British Journal of Dermatology, 155(2), 387–392. https://doi.org/10.1111/j.1365-2133.2006.07328.x

Chapter 6

Framing Attribute Levels in Case 2 Best-Worst Scaling: Do signs matter?

Soekhai V, Donkers B, van Kinschot CMJ, van Noord C and de Bekker-Grob EW

Submitted

Abstract

Objectives: Case 2 best-worst scaling (BWS-2) has become an increasingly popular method to elicit patient preferences in health and healthcare. However, this method is still in its infancy and estimation issues regarding mixing positive and negative attributes remain. Framing attributes positive or negative might solve this issue but leads to new challenges regarding preferences for gains and losses. The aim of this study was therefore to compare outcomes obtained from three different BWS-2 scenarios to study the impact of framing on outcomes.

Methods: Patients with Graves' disease completed an online survey, including eight BWS-2 tasks and – to test for convergent validity - 12 DCE tasks. For BWS-2, three different framing scenarios were used: regular, all positive or all negative. Patients were randomly assigned to one of these scenarios. Multinomial logit (MNL) and attribute-scale MNL (AS-MNL) were used to estimate preference weights and to calculate the relative importance of the attributes.

Results: A total of 192 patients were included for BWS-2 and DCE analyses. MNL analysis showed positive and negative BWS-2 outcomes differed in terms of the least preferred attribute levels as well as relative importance. For regular BWS-2 the ordering of attribute levels was not as theoretically expected. Results from the AS-MNL model showed that framing attributes negatively leads to attributes becoming more important, while framing attributes positively leads to attributes becoming less important.

Conclusions: This study showed that attribute framing in BWS-2 impacts outcomes and using regular BWS-2 leads to theoretically implausible outcomes. BWS-2 can be a useful method to elicit preferences but mixing positive and negative attributes should be avoided. Careful consideration about framing with regard to the decision context is required.

6.1 Introduction

Patient preferences have become more important in supporting medical decision-making at the individual and policy level in health and healthcare [1–3]. Studies by the Medical Device Innovation Consortium (MDIC) [4], Mahieu et al. [5] and Soekhai et al. [6] provide an overview of different stated preference methods to elicit preferences. A method that has become increasingly popular to elicit preferences in health and healthcare is best-worst scaling (BWS) [7,8]. BWS was introduced to obtain more preference information than a traditional discrete choice experiment (DCE) by asking individuals not only to select their best but also their worst option, without a large increase in the cognitive burden of the elicitation task [7]. For more information about DCEs see Train [9], Hensher et al. [10] and Soekhai et al. [11].

Louviere et al. [12] distinguish three types of BWS: object case (case 1 BWS) where attributes, profile case (case 2 BWS) where attribute levels, and multi-profile case (case 3 BWS) where profiles are selected as best and worst. For more details about BWS see Louviere et al. [12]. Noteworthy, case 2 BWS (hereafter: BWS-2) received much attention in the literature, as this method can uncover attribute level importance, can reduce cognitive burden of the elicitation task by focusing on one profile at a time and is relatively easy to design [13,14].

Several guidelines and best practices for conducting preference studies exist, in particular for DCEs [15,16]. Although BWS-2 is becoming more often used in health, the method is still in its infancy and several issues related to the design and analyses require further research [8]. One of these issues is the role of attribute level framing in BWS-2 choice tasks. Behavioral economic theory suggests that individuals cope differently when dealing with gains (e.g. increased life expectancy) or losess (e.g. more frequent side effects), with individuals placing more weight on losess than similar sized gains [17,18]. Currently, it is unknown what the impact of framing attributes differently (i.e. framed positively (a gain), negatively (a loss), or mixed (mix of gains and losses) is on BWS-2 outcomes. This is especially the case for BWS-2 since this is not a trade-off method like for example DCE but rather focuses on choices for attribute levels in isolation.

A second reason to study attribute framing in the BWS-2 context is that previous work suggests that BWS-2 with a mix of positive and negative attributes leads to dominance and therefore estimation problems [19]. To avoid this problem, attributes should be either framed all positively or all negatively. It is therefore important to understand the impact of different framings of BWS-2 attributes on BWS-2 outcomes.

Since there are to our knowledge no studies investigating framing effects in BWS-2, the main aim of this paper is to study the impact of attribute framing on BWS-2 estimates and the consequences of designing BWS studies with all attributes in either a positive or a negative frame. This will provide much needed guidance on designing BWS-2 tasks. In addition, we empirically compared BWS-2 to DCE preference weights to obtain insights into the convergent validity of the BWS-2 outcomes of the different attribute framings.

6.2 Methods

6.2.1 Study population

A sample of adult patients with morbus Graves' disease (hereafter: GD) was selected between August 2019 and July 2020. The rationale for using GD patients in this study was that there is insufficient data available on GD patients' treatment preferences. Respondents were recruited in multiple hospitals across the Netherlands, through press releases on websites of national endocrine organizations and patient organizations as well as through social media. Informed consent was obtained before the start of the survey. Patients who had a first diagnosis of GD or a recurrence in the previous year were included, while patients who were insufficiently fluent in the Dutch language were excluded. The study protocol was approved by the medical ethics committee from the Erasmus MC – University Medical Centre Rotterdam (MEC-2018-1665).

6.2.2 Attributes and attributes levels

Potentially relevant attributes and attribute levels related to GD treatment were selected using a multi-step approach for both BWS-2 as well as DCE. First, a literature search was conducted to identify attributes which were discussed with two medical researchers. Second, a focus group with fifteen GD patients was conducted to further elaborate on the attributes identified from the literature. Based on the literature search and focus group results, five attributes were included in the experiment: type of treatment, chance of being cured, chance of severe side effects, chance of permanent voice changes (treatment could lead to damage to vocal cord) and chance of hypocalcemia (treatment could lead to calcium deficiency). The attribute levels were based on information in the literature, followed by a consensus discussion by two medical researchers [20-22]. Table 1 presents the attributes and attribute levels for each BWS-2 scenario and DCE.

6.2.3 Design of BWS choice tasks

Since the aim of this study is to investigate framing effects of BWS-2 attributes on outcomes, three BWS-2 scenarios were developed that differed in the attribute framing. In the regular BWS-2 scenario, attributes in the choice tasks were included in their natural way, resulting in a mix of positive and negative attributes. To avoid comparisons of positive and negative attributes, in positive BWS-2 and negative BWS-2 all attributes were framed positively or negatively respectively (see Table 1). For each of the three BWS-2 scenarios an orthogonal

Table 1 – Attributes and levels for eliciting preferences with BWS-2 and DCE (including priors for DCE design)

Attributes	Levels			
Treatment type	**Medication** [Ref]	**Surgery** [-0.10,0.10][i] [0.98,0.28][ii]	**Radioactive Iodine** [-0.10,0.10][i] [-0.21,0.15][ii]	
Chance of being cured	**40%** [Ref]	**70%** [0.05,0.15][i] [0.71,0.28][ii]	**85%** [0.15,0.25][i] [1.09,0.23][ii]	**100%** [0.25,0.35][i] [1.30,0.36][ii]
Chance of severe side effects	**0%** [Ref]	**2%** [-0.15,-0.05][i] [-0.25,0.13][ii]	**5%** [-0.25,-0.15][i] [-0.37,0.14][ii]	**10%** [-0.35,-0.25][i] [-0.78,0.34][ii]
Chance of permanent voice changes	**0%** [Ref]	**5%** [-0.15,-0.05][i] [-0.50,0.16][ii]	**10%** [-0.25,-0.15][i] [-0.76,0.20][ii]	
Chance of hypocalcemia	**0%** [Ref]	**5%** [-0.15,-0.05][i] [-0.15,0.10][ii]	**10%** [-0.25,-0.15][i] [-0.22,0.11][ii]	**15%** [-0.35,-0.25][i] [-0.65,0.24][ii]

i uniformly distributed pilot prior: min-max
ii normally distributed post-pilot updated prior: mean, standard deviation

main effect plan (OMEP) experimental design was used. This type of design enables the independent estimation of preference weights for each attribute level [12]. Based on the number of attributes and levels, the OMEP indicated 16 choice tasks to be included in the survey [23]. Each BWS-2 scenario consisted of the same attributes and levels since the aim was to compare the different BWS-2 scenarios with each other (Table 1). In each BWS-2 scenario the attribute levels are identical but framed differently. This means that for positive BWS-2 the attributes chance of severe side effects, chance of permanent voice changes and chance of hypocalcemia were framed positively (e.g., chance of <u>no</u> severe side effects, levels 90%, 95%, 98%, 100%). Likewise, for negative BWS-2 the attribute chance of being cured was framed negatively (chance of <u>not</u> being cured, levels 0%, 15%, 30%, 60%). Attribute order was kept constant across all tasks.

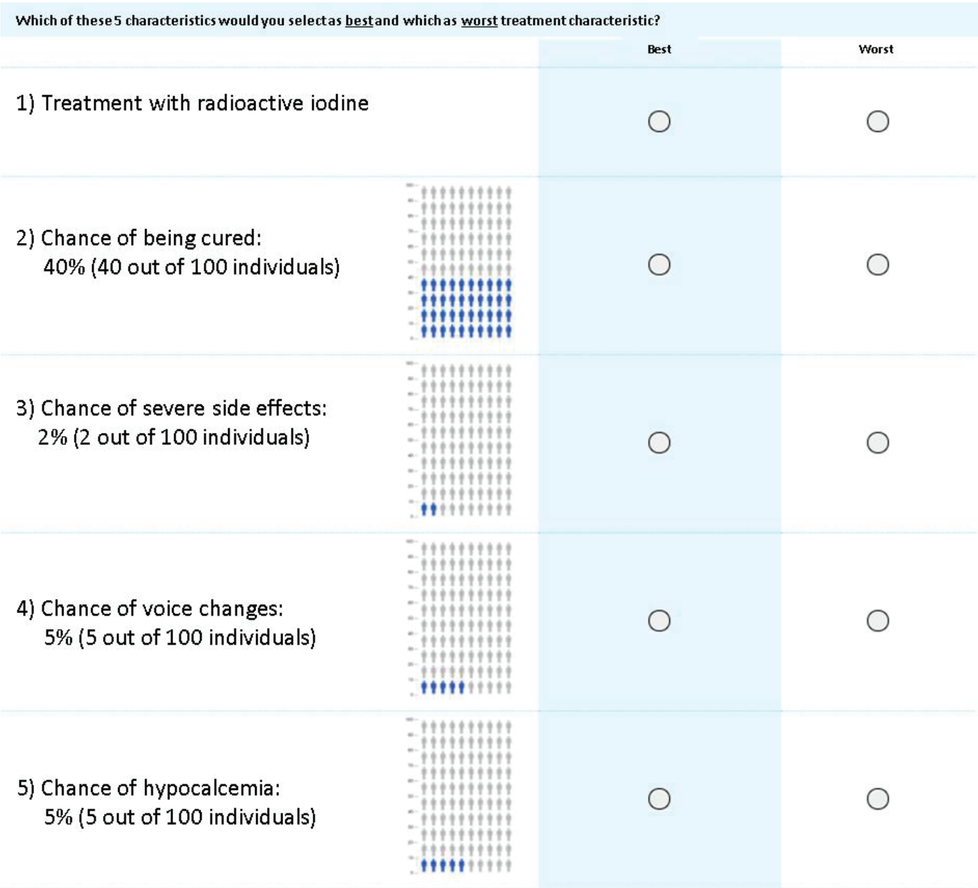
6.2.4 Design of DCE choice tasks

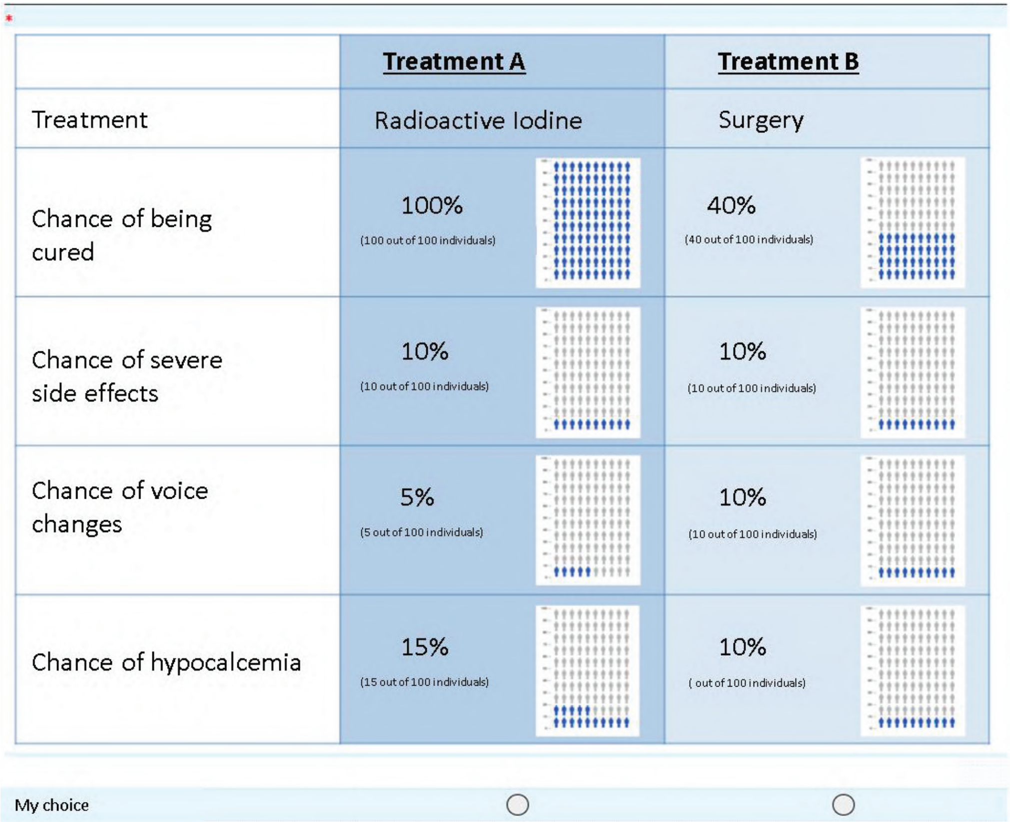
For the DCE, a Bayesian D-efficient design was generated in which the D-efficiency was maximized using Ngene software (Ngene, version 1.2.1) [16]. Pilot data from 32 patients were used to update priors and their distribution (see Table 1) as well as further optimization of the experimental design [10,15,16]. The generated design used for the survey included 48 choice tasks, which was blocked into four blocks with 12 choice tasks each to reduce cognitive burden for respondents. The alternatives in each choice task were unlabeled and the attribute order was kept constant across all tasks [24].

6.2.5 Survey design

The survey consisted of five sections: an introduction explaining the survey relevance and information about GD and treatment options, followed by background questions about sociodemographic characteristics and medical history, DCE tasks, BWS-2 tasks and final questions including a survey evaluation. Patients were randomly assigned to one of the three BWS-2 scenarios and had to complete both BWS-2 and DCE tasks. To reduce the cognitive burden, patients were requested to answer a completely random subset of 8 of the 16 BWS-2 tasks in the design. In the DCE, patients were asked about their preferences by choosing between two alternatives in 12 choice tasks, whereas in BWS-2 they had to select their best and worst attribute level. Before both the first BWS-2 and DCE choice task, a short introduction with an example choice task was presented to patients. In the final section of the survey respondents were asked to explicitly state which attribute they regarded as most and least important, as well as their overall evaluation of the survey. The validity of the survey was tested with five patients in a think-aloud format with direct verbal feedback to optimize the survey (i.e., better instructions and example tasks). The survey was designed using

LimeSurvey (LimeSurvey, version 2.06). A sample BWS-2 and DCE choice task is shown in Figure 1.





	Treatment A	Treatment B
Treatment	Radioactive Iodine	Surgery
Chance of being cured	100% (100 out of 100 individuals)	40% (40 out of 100 individuals)
Chance of severe side effects	10% (10 out of 100 individuals)	10% (10 out of 100 individuals)
Chance of voice changes	5% (5 out of 100 individuals)	10% (10 out of 100 individuals)
Chance of hypocalcemia	15% (15 out of 100 individuals)	10% (out of 100 individuals)
My choice	○	○

Figure 1 – Example regular BWS-2 and DCE choice tasks

6.2.6 Statistical analysis

The statistical analyses were performed using data from respondents who completed both BWS-2 and DCE tasks. Following guidance from the literature, statistical analyses were started by using multinomial logit modelling (MNL) [10]. Since the aim of this study is to investigate the impact of different attribute framings on BWS-2 outcomes, we also analyzed BWS-2 data using an MNL based econometric model allowing the utility scale of an attribute to differ depending on the framing used, which we will refer to in this study as the attribute-scale multinomial logit model (AS-MNL). Econometric models that allow for varying preferences between conditions have been used before by for example Dellaert et al. [25]. For BWS-2 estimations, scale differences between best and worst choices are allowed with the estimation of a best-worst scale parameter (beta_worst) in our model [26,27]. We aimed to include approximately 200 respondents for both BWS-2 (in total for the three scenarios) as well DCE to ensure sufficient statistical power for modelling [28-31].

Using MNL, the utility (U) of an alternative in both BWS-2 and DCE can be modeled as a linear function of the specific attributes and levels, with

$$U = \sum_{k=1}^{A} \sum_{j=1}^{J_k} \beta_{k,j} X_{k,j} + \varepsilon \qquad \text{eq. 1}$$

where there are A attributes with attribute *k* having J_k attribute levels, with $X_{k,j}$ equal to one if the attribute level j of an attribute k is available in the presented profile, $\beta_{k,j}$ are the utility parameters for the j^{th} levels of attribute k and ε being the random error term representing the unexplained part of utility. Where the MNL model assumes the same utility scale for each attribute, the AS-MNL model allows for differences in utility scales at the level of an attribute depending on the framing being used for that attribute. Scale parameters are introduced at the attribute level to allow for a shift in the importance of the attribute when it is framed different from the regular BWS-2 scenario. The resulting utility of an alternative is specified as:

$$U = \sum_{k=1}^{A} \lambda_k \sum_{j=1}^{J_k} \beta_{k,j} X_{k,j} + \varepsilon \qquad \text{eq. 2}$$

with λ_k being the scale parameter that is set to 1 for regular BWS-2 and is allowed to deviate from 1 when a positive (negative) attribute is framed negatively (positively). Additional scale parameters (positive and negative BWS-2) were estimated to capture the effect of a positive and negative framing. Both MNL and AS-MNL were programmed using R version 4.0.0

(Apollo package, version 0.0.1) to estimate the utilities for both the BWS-2 and DCE data [26,27]. For all three BWS-2 scenarios and DCE, treatment type medication was selected as reference level (fixed at zero) since this treatment type is considered the status quo.[viii] The DCE approach also requires a reference level within each attribute. In order to create a clear interpretation of attribute levels, for curation the least attractive attribute level and for the other attributes the most attractive attribute levels were selected as reference level. In that way for curation preference weights increase when the attribute level value increases, while for the other attributes the preference weights decrease with increasing attribute levels. To facilitate the comparison between BWS-2 and DCE, the utility levels relative to the corresponding attribute reference level were also estimated for BWS-2. Parameter estimates from the BWS-2 scenarios and DCE were plotted against each other and their fit was compared by looking at the R^2 values. Relative importance of attributes was calculated by looking at the maximum utility differences between two attribute levels within each specific attribute and compared between the three BWS-2 scenarios and DCE.

6.3 Results

In total 192 patients completed both the BWS-2 and DCE part of the survey. Responding patients were - as expected from the GD patient population - mostly female (between 91-95%) and the median age was between 45 and 47 years for the three BWS-2 scenarios and DCE. Most respondents completed positive BWS-2 (n=76), while n=63 completed negative BWS-2 and n=53 completed regular BWS-2. Of these, 71% (n=37), 61% (n=43) and 61% (n=38) patients experienced a first episode of GD looking at regular, positive and negative BWS-2, respectively. Between 45-52% of respondents in the BWS-2 subgroups completed a higher education. Within each subgroup, most patients were treated with medication in the

[viii] From a statistical perspective it does not matter which attribute level is selected as reference level.

last six months: 80% (n=42) in regular, 79% (n=60) in positive and 85% (n=53) in negative BWS-2. Overall, the sample characteristics show that the BWS-2 subgroups are quite similar to each other (Table 2).

Table 2 – Sample characteristics

Characteristic	Regular BWS-2	Positive BWS-2	Negative BWS- 2	DCE
Respondents	53	76	63	192
Sex				
Female	48 (92%)	69 (91%)	59 (95%)	178 (93%)
Age (years)				
Median (range)	46 (20-73)	47 (20-87)	45 (19-62)	46 (19-87)
Highest level of education				
Elementary	0 (0%)	2 (3%)	0 (0%)	2 (1%)
Secondary	15 (28%)	11 (14%)	10 (16%)	36 (19%)
Vocational	11 (21%)	22 (29%)	17 (27%)	50 (26%)
Higher	24 (45%)	38 (50%)	33 (52%)	95 (49%)
Unknown	0 (0%)	2 (3%)	0 (0%)	2 (1%)
No Answer	3 (6%)	1 (1%)	3 (5%)	7 (4%)
Graves' disease				
First episode	37 (71%)	43 (61%)	38 (61%)	122 (64%)
Recurrent disease	13 (25%)	28 (38%)	22 (36%)	63 (33%)
Unknown	2 (4%)	1 (1%)	1 (3%)	6 (3%)
Previous treatment(s) in last 6 months				
Medication	42 (80%)	60 (79%)	53 (85%)	155 (81%)
Surgery	2 (4%)	5 (7%)	8 (14%)	15 (8%)
Radioactive Iodine	2 (4%)	3 (5%)	1 (3%)	7 (4%)
No treatment	5 (11%)	9 (12%)	4 (7%)	17 (9%)

Table 3 presents the estimated preference weights for all BWS-2 scenarios as well as DCE. The estimated preference weights indicate that for regular BWS-2 85% and 100%

chance of being cured were preferred most, while a 10% chance of side effects and a 15% chance of hypocalcemia were preferred least. For positive BWS-2 most preferred attribute levels were 100% chance of being cured and 0% chance of severe side effects, while least preferred attribute levels were 40% chance of curation and treatment with radioactive iodine. Looking at the estimated preference weights for the attribute levels of chance of severe side effects, statistically significant differences between positive and regular BWS-2 were found for all attribute levels (t-test, $p<0.05$). Additionally, this was the case for the attribute levels 10% chance of permanent voice changes and 15% chance of hypocalcemia. In the case of negative BWS-2 the most preferred attribute levels were a 100% chance of being cured and 0% chance of severe side effects, while attribute levels 40% and 70% chance of being cured were found to be the least preferred. Comparing the estimates for the chance of being cured attribute levels between negative and regular BWS-2, the estimates were statistically significantly different from each other (t-test, $p<0.05$) for all attribute levels, except for the attribute level 100% chance of being cured. For DCE it was found that respondents preferred a treatment that consisted of medication with the highest chance of being cured and the lowest chances of severe side effects, permanent voice changes and hypocalcemia.

Focusing on the magnitudes of the estimated preference weights for each attribute level for all BWS-2 scenarios and DCE, the more attractive curation levels were more preferred compared to less attractive levels. An important exception is the utility of an 85% chance of being cured for regular BWS-2, which was larger than that of a 100% chance of being cured, although the difference was not significant. Table 3 also showed that for the attributes chance of severe side effects, chance of permanent voice changes and chance of hypocalcemia less attractive levels were less preferred except for 10% chance of hypocalcemia within negative BWS-2.

Table 3 – Multinomial logit results for BWS-2 and DCE

	Regular BWS-2		Positive BWS-2		Negative BWS-2		DCE	
	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se
Treatment type								
Medication	REF	-	REF	-	REF	-	REF	-
Surgery	-2.07**	0.47	-1.48**	0.29	-1.45**	0.30	-1.55**	0.15
Radioactive Iodine	-2.60**	0.53	-2.28**	0.44	-1.97**	0.36	-1.70**	0.16
Chance of being cured (%)								
40	REF	-	REF	-	REF	-	REF	-
70	2.88**	0.31	1.03**	0.32	0.42*	0.23	1.40**	0.13
85	4.23**	0.34	2.19**	0.51	1.22**	0.42	1.83**	0.16
100	4.19**	0.44	5.13**	0.55	4.41**	0.44	2.48**	0.20
Chance of severe side effects (%)								
0	REF	-	REF	-	REF	-	REF	-
2	-1.86**	0.47	-0.91**	0.21	-1.75**	0.26	-0.19**	0.09
5	-2.32**	0.61	-1.10**	0.26	-2.36**	0.24	-0.82**	0.14
10	-3.22**	0.76	-1.87**	0.3	-2.65**	0.30	-1.31**	0.13
Chance of permanent voice changes (%)								
0	REF	-	REF	-	REF	-	REF	-
5	-0.92**	0.32	-0.80**	0.17	-1.09**	0.17	-0.23**	0.08
10	-1.70**	0.39	-1.19**	0.23	-1.44**	0.20	-0.34**	0.11
Chance of hypocalcemia (%)								
0	REF	-	REF	-	REF	-	REF	-
5	-0.82**	0.33	-0.43**	0.18	-0.96**	0.21	-0.22**	0.08
10	-1.46**	0.37	-0.62**	0.2	-1.45**	0.27	-0.22**	0.08
15	-2.39**	0.46	-0.95**	0.21	-1.43**	0.25	-0.78**	0.11
Reference levels[i]								
Treatment type: Medication	0	-	0	-	0	-	0	-
Chance of being cured (%): 40	-2.46**	0.40	-3.10**	0.45	-3.19**	0.45	0	-
Chance of severe side effects (%): 0	-0.61	0.39	1.06**	0.30	0.14	0.31	0	-
Chance of permanent voice changes (%): 0	-0.84**	0.30	-0.28	0.25	-0.53*	0.29	0	-
Chance of hypocalcemia (%): 0	-0.76**	0.26	-0.69**	0.24	-0.44	0.33	0	-
beta_worst[ii]	0.95	0.24	1.06	0.19	1.30	0.29	-	-
Sample size	53		76		63		192	
Log likelihood	-843.96		-1372.06		-1131.15		-1363.10	

* Significant at 10%
** Significant at 5%
[i] β estimates attribute levels estimated as additional utility or disutility compared to reference level
[ii] beta_worst parameter allows for scale differences between best and worst choices (hypothesis testing beta_worst = 1 showed no statistically significant outcomes)

The overall relative importance for all attributes is illustrated in Figure 2. For regular BWS-2 curation had the highest relative importance, followed by severe side effects, treatment type, hypocalcemia and permanent voice changes. The relative importance values for positive BWS-2 indicate that curation had the highest relative importance, followed by treatment type, side effects, permanent voice changes and hypocalcemia. The relative importance values from the DCE show a similar pattern, with the highest value for curation, followed by treatment type, severe side effects, hypocalcemia and permanent voice changes. Negative BWS-2 values follow a somewhat different pattern, with curation having the highest importance and permanent voice changes and hypocalcemia both low in importance. Based on theory we would expect attributes that are framed negatively (i.e., curation) becoming more important, while attributes that are framed positively (i.e., side effects, hypocalcemia and voice changes) becoming less important. The bar plots in Figure 2 indeed show that comparing the relative importance for curation in regular and negative BWS-2, this value increases in the negative BWS-2 scenario. The relative importance values for attributes side effects, hypocalcemia and voice changes decrease when comparing regular and positive BWS-2.

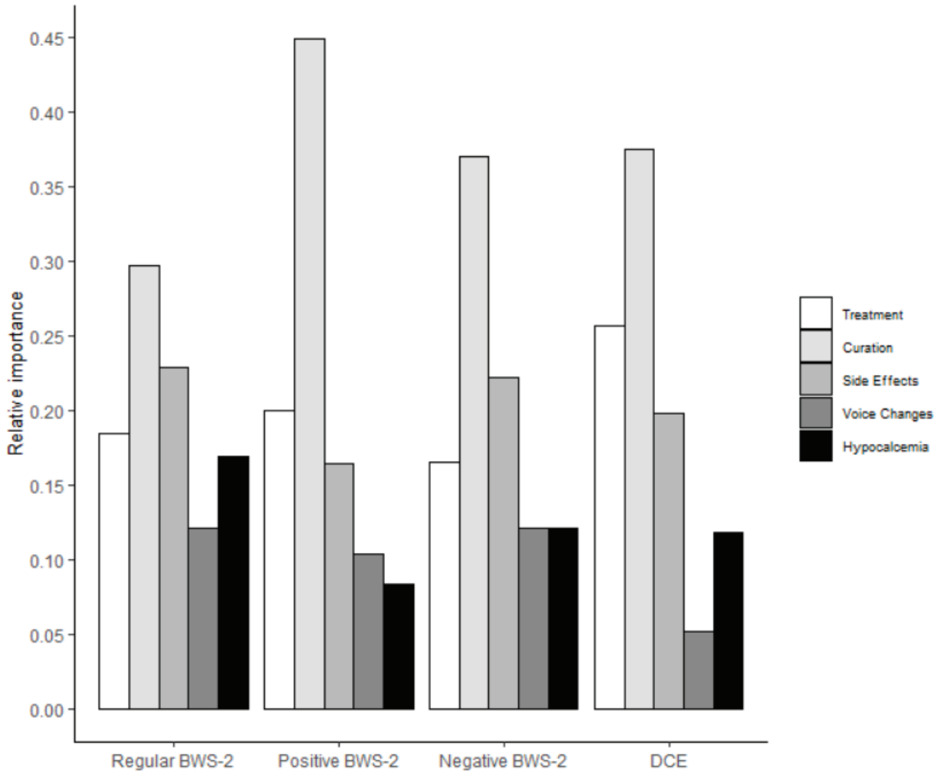


Figure 2 – Relative importance of attributes for BWS-2 and DCE

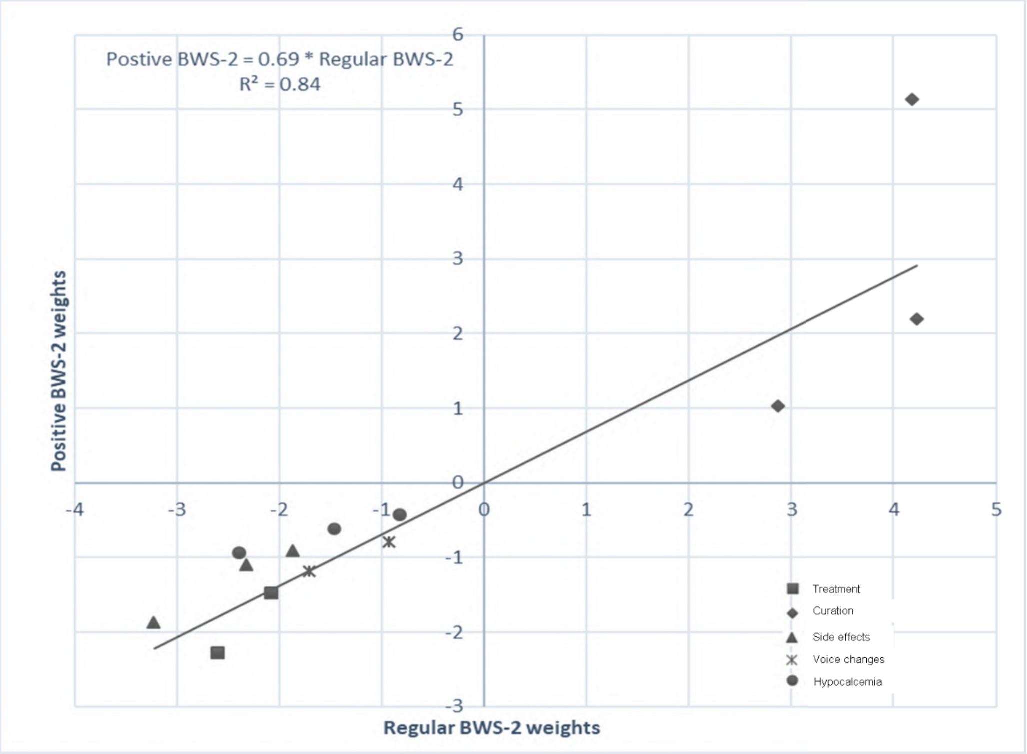
Table 4 presents the AS-MNL parameters for each attribute, which capture the impact of the change in attribute framing from the regular framing to a positive or negative framing. The results showed that framing the chance of being cured negatively, results in a scale parameter larger than 1, i.e., exp (0.96), indicating that this attribute becomes more important. The scale parameters for attributes chance of severe side effects, permanent voice changes and hypocalcemia indicate that framing these attributes positively will lead to a scale parameter smaller than 1 (taking the exponent of the negative values), meaning these attributes become less important.

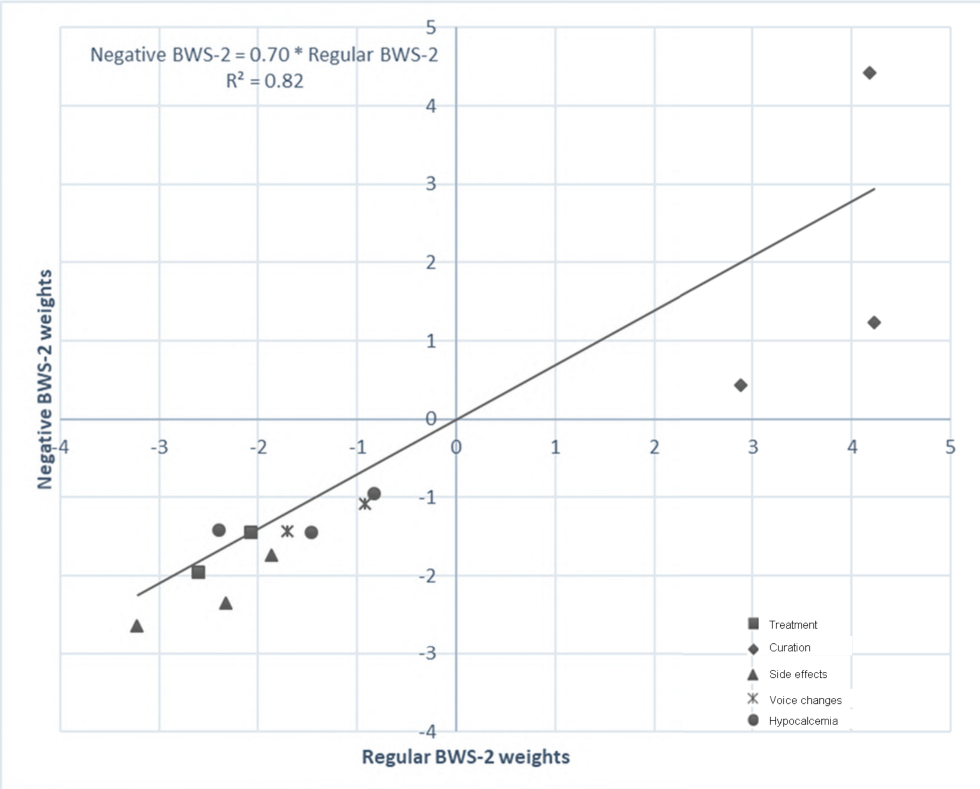
Table 4 – Attribute-scale adjusted multinomial logit parameters for pooled BWS-2 data

	Pooled BWS-2	
	Estimate	**Rob. Se**
Logarithm of scale parameters (log(lambda_k))[i]		
Chance of being cured	0.96**	0.16
Chance of severe side effects	-2.33**	0.39
Chance of permanent voice changes	-0.78**	0.13
Chance of hypocalcemia	-0.70**	0.13
Positive BWS-2	0.34**	0.15
Negative BWS-2	-0.45**	0.14
Sample size	192	
Log likelihood	-3528.47	

** Significant at 5%
[i] scale parameters that capture the effect of scale differences between positive and negative BWS-2

Comparing the estimated preference weights from regular BWS-2 to positive BWS-2 and negative BWS-2 directly (without the reference levels), the regression plot for positive BWS-2 indicated a better fit (R^2 = 0.84) compared to negative BWS-2 (R^2 = 0.82) (Figure 3). Additionally, the fit for DCE is overall the best with an R^2 of 0.90. Furthermore, for all plots the slope coefficient of the best fit line is smaller than one, suggesting that the positive BWS-2, negative BWS-2 and DCE weights are overall smaller than the regular BWS-2 weights. In these plots, the utility estimate for 100% chance of being cured is consistently and substantially higher for all other methods than what is predicted based on the regular BWS-2 based regression line.





Negative BWS-2 = 0.70 * Regular BWS-2
R² = 0.82
Negative BWS-2 weights
Regular BWS-2 weights
5
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
5
Treatment
Curation
Side effects
Voice changes
Hypocalcemia

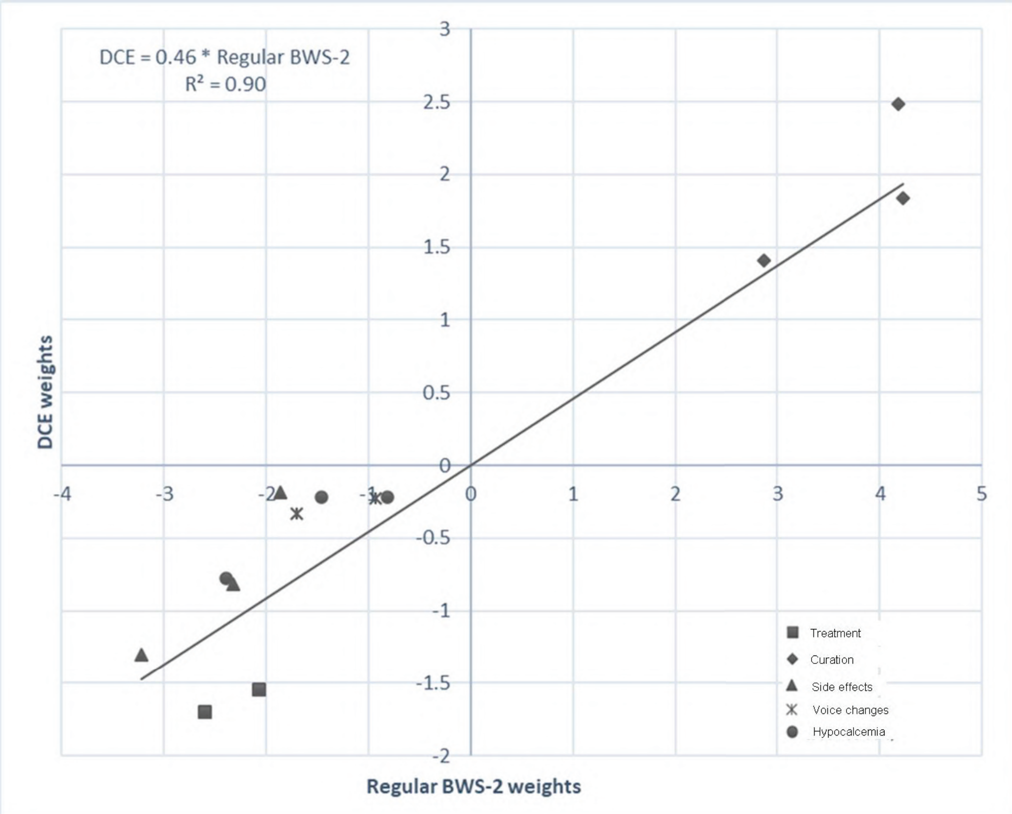


Figure 3 – Comparison between positive BWS-2, negative BWS-2, DCE and regular BWS-2 preference weights

6.4 Discussion

In this study we investigated the impact of attribute framing on BWS-2 estimates. We conclude that comparing positively and negatively framed BWS-2, a different preference pattern for the least preferred attribute level was found. The ordering of each attribute level was overall similar across methods, except for the highest chance of being cured in regular BWS-2 and the highest chance of hypocalcemia in negative BWS-2.

Although this is the first study comparing the impact of positive and negative framing on BWS-2 outcomes, there are several stated preference studies focusing on the impact of attribute framing. A study by Howard & Salkeld [32] showed that attribute framing in a DCE impacts the calculation of marginal rate of substitution (MRS) values. The authors also stated that using a positive or negative frame impacts the relative importance of attributes. Veldwijk et al. [33] also found similar results in their study about framing effects of risk attributes in a DCE. These conclusions are in line with our study, in which we found differences between relative importance values between positive and negative BWS-2. In this study we introduced the AS-MNL model to model differences between the framing scenarios that affect attribute importance. This model showed that framing attributes negatively leads to attributes becoming more important, while framing attributes positively leads to attributes becoming less important.

One of the main findings from the BWS-2 results was the fact that for regular BWS-2 more attractive levels of chance of being cured did not lead to these levels being more preferred. This might be caused by the mix of positive (cured) and negative (side effects, permanent voice changes and hypocalcemia) attributes that leads to estimation problems with BWS-2. Specifically, regular BWS-2 is not able to distinguish between the most attractive attribute levels when a single positive attribute is included as the attractive levels are selected as best very often – reaching a ceiling where the method cannot differentiate between the levels. Descriptives from our BWS-2 choice data for example showed that individuals selected both the most and second most attractive curation level as best in 88% and 83% of the choice tasks when this level was included, respectively. For positive (76% and 24%) and negative (81% and 24%) BWS-2 these percentages are much more different from each other, suggesting that a better distinction can be made between these attribute levels based on the data from these scenarios. An important underlying reason why this occurs in BWS-2 with mixed positive and negative attributes is that no direct comparisons of the levels within an attribute are ever made and identification needs to be obtained from comparisons of the levels of different attributes. The resulting identification problems are studied in more detail by

Soekhai et al. [19], where the authors advise to either frame all attributes positively or negatively when dealing with a mix of positive and negative attributes. The results from this study are in line with the results of Soekhai et al. [19] since framing all attributes positively in positive BWS-2 leads to outcomes consistent with our expectations while the results of regular BWS-2 lack face validity.

Based on behavioral economic theory it is known that individuals place more weight on avoiding losses than acquiring equivalent gains [17,18]. Therefore, we would expect negative attributes that are framed positively to become less important, meaning closer to zero. The AS-MNL model was used to gain insights into the effect of changes in framing on attribute importance. The results confirmed the theory-based predictions that attributes become less important when framed positively and more important when framed negatively.

The preference weight patterns between regular BWS-2 and DCE showed a comparable trend, which is an indication of the convergent validity of our regular BWS-2 outcomes, although – unlike DCE – BWS-2 did not identify a difference in preferences between an 85% and a 100% chance of being cured. The general consistency of preference weights is in agreement with previous studies comparing BWS-2 and DCE preference weights. Van Dijk et al. [13], Potoglou et al. [34] and Severin et al. [35] showed comparable patterns in attribute weights. The estimates, fit and slope (regular BWS-2 weights are roughly two times larger as DCE weights) of the best fit line from the direct comparison between regular BWS-2 and DCE weights in the scatterplot were overall in line with the study from van Dijk et al. [13]. Small differences between the latter study and our study might be related to differences in the health decision context.

A strength of this study is that it is the first study investigating the effects of attribute framing in BWS-2 and therefore filling a gap in (health) preference literature. As mentioned before, there are several DCE studies investigating the impact of attribute framing. However, to our knowledge, similar studies have not been conducted for BWS-2. We believe this is especially important for BWS-2 since individuals make choices on an attribute level instead of the profile level, omitting direct trade-offs across attribute levels, which are required in for example a DCE but also in real-life decisions. Another strength of this study is that the attributes and levels were selected using a multi-step approach, consisting of a literature search and focus groups, including pretesting the survey before data collection was started. By combining the designs of both BWS-2 and DCE, with a qualitative approach and pretesting, the content validity of the study was increased [36].

Although we followed guidelines and best practices in developing BWS-2 and DCE, this study has limitations [10,12]. Importantly, we did not randomize the BWS-2 and DCE part of the experiment. In our survey patients always first completed DCE tasks before starting with the BWS-2 tasks. It could therefore be that BWS-2 responses might be influenced by DCE responses. However, pretesting showed that patients had more difficulty understanding the concept of BWS-2 compared to DCE. In order to get a proper introduction to the stated choice experiment concept, we decided to always place DCE before BWS-2.

6.5 Conclusion

This study showed that different attribute framings in BWS-2 lead to different outcomes when looking at preference weights and attributes' relative importance. We showed that positive BWS-2 framing outcomes were more similar to DCE outcomes than negative BWS-2 framing outcomes. Our results also indicated that mixing positive and negative attributes in BWS-2 can lead to undesirable outcomes. Additionally, we showed that when framing attributes negatively in BWS-2, these attributes become more important for respondents. When framing attributes positively, these attributes become less important. This will impact estimation outcomes and therefore conclusions based on the specific framing being used. Hence, we advise to carefully consider how to frame attributes in BWS-2 surveys. It will be important that the frame matches the decision environment that the respondent is facing as well as the fact that attributes and levels should be always presented in such a way that they are relevant and relatively easy to interpret. Qualitative work and pretesting will be important to validate this. BWS-2 can be a useful method to elicit preferences but mixing positive and negative attributes should be avoided as it might cause ceiling effects in preference estimates. Careful consideration about framing attributes either positively or negatively will be important to get useful outcomes to support medical decision-making.

References

1. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health - A checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Heal [Internet]. 2011 [cited 2018 Jun 23];14:403–13. Available from: https://www.ispor.org/TaskForces/documents/ISPOR-CA-in-Health-TF-Report-Checklist.pdf

2. European Medicines Agency. Report of the CHMP working group on benefit-risk assessment models and methods. Reproduction [Internet]. 2007;16. Available from: Doc. Ref. EMEA/CHMP/15404/2007

3. Food and Drug Administration (FDA). Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling [Internet]. 2016. Available from: http://www.regulations.gov.

4. MDIC. Medical Device Innovation Consortium Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Med Device Innov Consort [Internet]. 2015;

5. Mahieu P-A, Andersson H, Beaumais O, Crastes dit Sourd R, Hess S, Wolff F-C. Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. Rev Agric Food Environ Stud. 2017;98:201–20.

6. Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. Drug Discov. Today. Elsevier Ltd; 2019. p. 1324–31.

7. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. J Health Econ [Internet]. 2007 [cited 2019 Jan 31];26:171–89. Available from: https://ac.els-cdn.com/S016762960600049X/1-s2.0-S016762960600049X-main.pdf?_tid=0e561dd4-cf96-41d5-8a9c-2debef6ff88f&acdnat=1548942271_9ae125c7d84bfa06fa009d79de40d56d

8. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ Rev [Internet]. SpringerOpen; 2016 [cited 2019 Jan 31];6:1–14. Available from: http://healtheconomicsreview.springeropen.com/articles/10.1186/s13561-015-0079-x

9. Train, K.E., 2009. Discrete Choice Methods with Simulation, Second Edition. Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

10. Hensher DA, Rose JM, Greene WH. Applied choice analysis: A primer [Internet]. Appl. Choice Anal. A Prim. Cambridge University Press; 2015 [cited 2020 Sep 29]. Available from: /core/books/applied-choice-analysis/7554FC97EDA7C61D29BE3DA7FF5CBD9B

11. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future [Internet]. Pharmacoeconomics. Springer International Publishing; 2019 [cited 2020 Sep 15]. p. 201–26. Available from: http://go.ncsu.edu/ellis

12. Louviere JJ, Flynn TN, Marley AAJ. Best-worst scaling: Theory, methods and applications [Internet]. Best-Worst Scaling Theory, Methods Appl. Cambridge: Cambridge University Press; 2015 [cited 2019 Jan 31]. Available from: http://ebooks.cambridge.org/ref/id/CBO9781107337855

13. van Dijk JD, Groothuis-Oudshoorn CGM, Marshall DA, IJzerman MJ. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. Value Heal [Internet]. Elsevier; 2016;19:316–22. Available from: http://dx.doi.org/10.1016/j.jval.2015.12.020

14. 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. Med Decis Mak [Internet]. 2014 [cited 2019 Jan 31];34:638–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24713695

15. Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Heal. Elsevier Ltd; 2016;19:300–15.

16. Johnson FR, 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 Heal [Internet]. 2013 [cited 2018 Jun 23];16:3–13. Available from: https://www.ispor.org/ValueInHealth/ShowValueInHealth.aspx?issue=3551E3F3-83A8-4895-9415-5272AF29A9DF

17. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Decis Probab Util [Internet]. Cambridge University Press; 1988 [cited 2019 Nov 29]. p. 183–214. Available from: https://www.cambridge.org/core/product/identifier/CBO9780511609220A023/type/book_part

18. Tversky A, Kahneman D. The Framing of Decisions and the Psychology of Choice [Internet]. 1981. Available from: www.sciencemag.org

19. Soekhai V, Donkers B, de Bekker-Grob E. PNS295 Best Worst Scaling: For Good Or For Bad But Not For

Both. Value Heal. Elsevier BV; 2019;22:S813.

20. Burch HB, Cooper DS. Management of graves disease a review. JAMA - J Am Med Assoc. 2015;314:2544–54.

21. Cooper DS. Hyperthyroidism. Lancet Seminar. 2003;362:459–68.

22. Smith, Terry J., Hegedüs L. Graves' Disease. N Engl J Med. 2016;

23. Hahn, G., 1966. A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. General Electric Research and Development Center, Schenectady New York.

24. De Bekker-Grob EW, Hol L, Donkers B, Van Dam L, Habbema JDF, Van Leerdam ME, et al. Labeled versus unlabeled discrete choice experiments in health economics: An application to colorectal cancer screening. Value Heal. 2010;13:315–23.

25. Dellaert BGC, Donkers B, Van Soest A. Complexity effects in choice experiment-based models. J Mark Res. 2012;49:424–34.

26. Hess S, Palma D. Apollo : a flexible , powerful and customisable freeware package for choice model estimation and application Stephane Hess & David Palma Choice Modelling Centre University of Leeds. Apollo user Man. 2019

27. Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. J Choice Model. 2019;32:1–43.

28. Lancsar E, Louviere J. PRACTICAL APPLICATION Conducting Discrete Choice Experiments to Inform Healthcare Decision Making A User's Guide. Pharmacoeconomics [Internet]. 2008 [cited 2018 Jul 13];26:661–77. Available from: https://link.springer.com/content/pdf/10.2165%2F00019053-200826080-00004.pdf

29. Noordman BJ, de Bekker-Grob EW, Coene PPLO, van der Harst E, Lagarde SM, Shapiro J, et al. Patients' preferences for treatment after neoadjuvant chemoradiotherapy for oesophageal cancer. Br J Surg. 2018;105:1630–8.

30. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. Patient [Internet]. Springer International Publishing; 2015 [cited 2018 Jun 27];8:373–84. Available from: http://link.springer.com/10.1007/s40271-015-0118-z

31. Louviere JJ, Flynn TN, Marley AAJ. Best-worst scaling: Theory, methods and applications [Internet]. Best-Worst Scaling Theory, Methods Appl. 2015 [cited 2018 Jun 22]. Available from: http://admin.cambridge.org/se/academic/subjects/economics/econometrics-statistics-and-mathematical-

economics/best-worst-scaling-theory-methods-and-applications#rKSkSCVw0d1z2GHT.97
32. Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. Value Heal [Internet]. International Society for Pharmacoeconomics and Outcomes Research (ISPOR); 2009;12:354–63. Available from: http://dx.doi.org/10.1111/j.1524-4733.2008.00417.x
33. Veldwijk J, Essers BAB, Lambooij MS, Dirksen CD, Smit HA, De Wit GA. Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? Value Heal [Internet]. Elsevier; 2016;19:202–9. Available from: http://dx.doi.org/10.1016/j.jval.2015.11.004
34. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc Sci Med. England; 2011;72:1717–27.
35. 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. Eur J Hum Genet [Internet]. 2013 [cited 2019 May 19];21:1202–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23486538
36. Vass C, Rigby D, Payne K. The Role of Qualitative Research Methods in Discrete Choice Experiments. Med Decis Making [Internet]. SAGE Publications; 2017 [cited 2018 Jul 25];37:298–313. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28061040

Chapter 7

Implicit versus Explicit Reference Points in Case 2 Best-Worst Scaling Tasks

Soekhai V, Donkers B and de Bekker-Grob EW

Submitted

Abstract

In this study we investigated the benefits of BWS-2R, which includes explicit reference points in BWS-2 to overcome results that are driven by (unobserved) differences in reference points between individuals. We analytically showed that BWS-2R should reduce noise in the inferred preferences as regular BWS-2 preferences are confounded with the reference points. Therefore, BWS-2R should lead to a more accurate representation of preferences. Our empirical study results showed statistically significant differences between estimated preference weights BWS-2 and BWS-2R. Also, statistically significant differences in RI scores between BWS-2R and BWS-2 were found. Our results also showed no difference in perceived difficulty between BWS-2R and BWS-2, with a larger proportion of respondents that completed BWS-2 preferring BWS-2R than the other way around. Hence, we advise using BWS-2R when aiming to conduct a BWS-2 experiments for preference research in health economics.

7.1 Introduction

The use of patient preferences in supporting medical decision-making has become more important in recent years.[1,2] Studies by for example the Medical Device Innovation Consortium (MDIC)[3], Mahieu et al.[4] and Soekhai et al.[5] provided an overview of different stated preference methods to elicit these preferences. Best-worst scaling (BWS) has become an increasingly popular method to elicit preferences in health and healthcare.[6,7] The introduction of BWS originated from the intent to obtain more preference information than from a traditional discrete choice experiment (DCE) by asking individuals to also select their least preferred option, without increasing the cognitive burden.[8,9] In health and healthcare BWS is often used for health state and medical treatment valuation.[7]

There are three types of BWS: object case (case 1 BWS) where attributes, profile case (case 2 BWS) where attribute levels, and multi-profile case (case 3 BWS) where profiles are selected as best and worst, see Appendix A and Louviere et al. for more details.[8] Case 2 BWS (hereafter: BWS-2) is an interesting method which reduces cognitive burden (relative to case 3 BWS or DCE) by focusing on a single profile at a time, uncovers attribute level importance and is relatively easy to design.[10,11] While much is known about case 3 BWS due to its similarities to DCEs, with the increasingly popular case 2 BWS there are still several issues relating to its design and analysis that require further investigation.

One of these issues is the interpretation of BWS-2 choice tasks. In traditional BWS-2 individuals are presented with multiple attribute levels and are asked to select their best and worst attribute level. For example, when an individual needs to decide whether “drug can be combined with alcohol” is best, s/he needs to imagine what s/he receives when not selecting this attribute level. In this case it will be clear: “drug cannot be combined with alcohol”. For other types of attributes this might be not that simple. When an individual for example needs to decide if attribute “60% effectiveness” is best, it is not clear what “60% effectiveness” is compared to (as the analyst cannot look into the individual's head): to other attribute levels in the choice task only and/or another point of comparison? Although not presented explicitly, it is likely that respondents evaluate attribute levels while having a certain point of comparison in mind. In this case it could be that an individual believes a typical treatment has a "40% effectiveness" to which the "60% effectiveness" is compared to. This point of comparison could also be another value, e.g. full effectiveness or no effectiveness. Moreover, for other attributes other levels will likely be in place. In this study we will use a broad interpretation of this point of comparison and will refer to this as the reference point.

The role of reference points and their relation to choices in BWS-2 is not well-defined. More specifically, although reference points are likely to influence the utilities and therefore the choices respondents make, they are not included in the choice task nor in the analysis.[13] In this study we therefore introduce a new BWS-2 approach which includes explicit reference points for each attribute level to ensure a known and stable point of comparison for all respondents.

The aim of this paper is to introduce a new BWS-2 approach with explicit reference points and to investigate whether this new approach leads to a more accurate analysis of preferences compared to BWS-2. This study will be an important step to further advance our understanding of BWS-2 choice experiments as well as further develop best practices.

7.2. Reference points in BWS-2

7.2.1 The role of reference points in traditional BWS-2

In traditional case 2 BWS (BWS-2) respondents are presented with multiple attribute levels and are asked to select the "best" and "worst" option.[14] Choices for "best" and "worst" in BWS-2 are often analyzed within the Random Utility Theory (RUT) framework.[14] This theoretical framework underpins the choice processes used to analyze BWS-2 choices.[15–17]

According to the RUT framework in BWS-2, respondents obtain a certain amount of utility (U) from each level of each attribute presented in the choice task. Regarding the best or worst question, a respondent selects the option that provides the highest or lowest utility, respectively.

When a respondent selects attribute 1, level 3 as best, and hence better than attribute 2, level 3, in a BWS-2 task, the utility of attribute 1, level 3 is larger than the utility of attribute 2, level 3:

$$U_{1,3} > U_{2,3} \qquad \text{eq. 1}$$

As mentioned in the introduction, it is likely that respondents evaluate attribute levels in BWS-2 tasks while having a certain (implicit) reference point in mind, for example: When respondents expect the effectiveness of a treatment to be 60%, while the true effectiveness is 40%. Table 1 shows a hypothetical example with 2 attributes, each with three levels. Each attribute level ($A_{1,1}$ meaning attribute 1, level 1) provides the respondent with a certain amount of utility on the utility scale (Figure 1). Without taking reference points into account, both respondents 1 and 2 will select $A_{1,3}$ over for example $A_{2,3}$ as best since this level provides them the highest utility.

Table 1 Example attributes and levels for BWS-2R

Attribute	Levels			
Attribute 1 (A_1)		$A_{1,1}$	$A_{1,2}$	$A_{1,3}$
e.g. effectiveness of treatment	e.g.	20%	40%	60%
Attribute 2 (A_2)		$A_{2,1}$	$A_{2,2}$	$A_{2,3}$
e.g. frequency of treatment	e.g.,	1/year	2/year	3/year

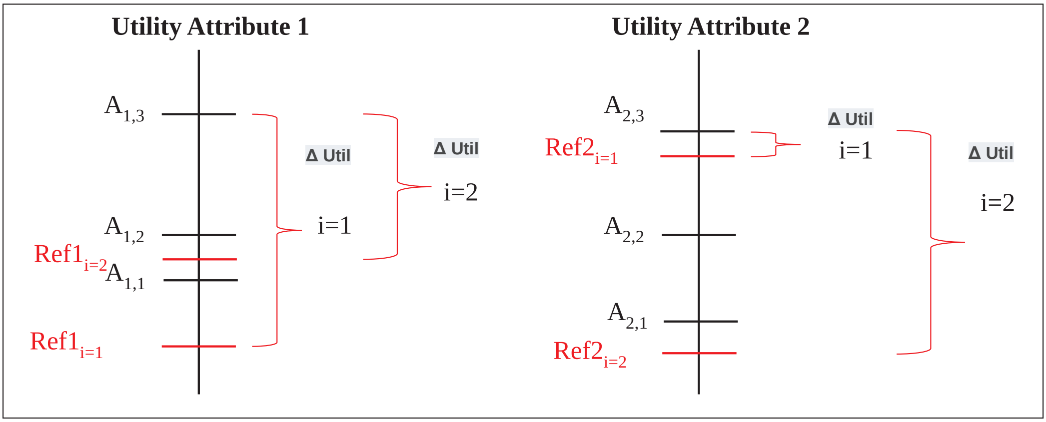


Figure 1 Utility levels of attribute levels with different individual specific reference points

However, taking into account the utility of reference points (Ref1 and Ref2, for attributes 1 and 2 respectively), the attribute level utilities will be influenced by these reference point utilities. These utilities represent the utilities of the points of comparison the attribute levels are compared to: When the utility of an attribute level is larger than the reference point utility, the utility change from reference point to attribute level will be positive. On the other hand, when the utility of an attribute level is smaller than the reference point utility, the utility change from reference point to attribute level will be negative. We assume that choices in BWS-2 are based on the utility differences between the attribute level and the reference point. Hence, respondent 1 will select $A_{1,3}$ as best when the difference in utilities between $A_{1,3}$ and Ref_1 is greater than the utility differences between the other attributes and their reference points. More formally, $A_{1,3}$ will be selected as best over $A_{2,3}$ when:

$$(Best = A_{1,3}) = ((U_{1,3} - U_{Ref1}) > (U_{2,3} - U_{Ref2})) \qquad \text{eq. 2}$$

This shows that reference points have an impact on the choice outcomes and these points potentially differ across respondents.[ix] When respondents are being asked to select the best attribute level, they will evaluate attribute levels relative to a certain reference point. When reference points are not presented explicitly, individuals' implicit reference points will affect the utility of the attribute level and therefore their best and worst choices.

Keeping eq. 2 in mind, when for example the reference point utility (U_{Ref}) for an individual is relatively high, attribute levels will be selected as best less often compared to when this reference point utility is relatively low. Therefore, taking or not taking reference points into account in BWS-2 may lead to different choices. Hence, differences in choices can result from two possible sources. First, because of differences in reference points. Second, because of actual differences in preferences. Since BWS is used to measure preferences, this method aims to capture actual differences in preferences. The inferred preferences should hence not be driven by reference point differences. Not accounting for this will lead to preference measurements that are confounded by reference points levels since choices in BWS-2 are not only drive by an individual’s preferences but also dependent on the individual’s reference points.

It should also be noted that besides differences in reference points between individuals (reference point heterogeneity), it could also be that reference points are not easy to imagine for individuals (reference point ambiguity). An individual for example finds it easier to imagine a reference point for the number of days to wait before a test result than a reference point for the effectiveness of a drug (s)he never used. In this latter example the individual lacks knowledge and context, making it more difficult to imagine a realistic reference point. In these situations, analyses could indicate that attributes with a more ambiguous reference

[ix] Results from our empirical study showed that for the attribute effectiveness (percentage) in BWS-2, the individual reference points ranged from 2.44 to 100.00, with a mean of 49.05 and a standard deviation of 26.92.

point have less impact while in real life they could play an important role in the choices the individual makes.

7.2.2 Introducing BWS-2R: BWS-2 with reference points

Since reference points play an important, yet implicit role in BWS-2, we propose a new BWS type which we name BWS-2R. In BWS-2R we provide explicit reference points for each attribute in the choice task. As illustrated in eq. 2, the selection of attribute levels as best (or worst) depends on the utility difference between attribute level and reference point utility. In the example in Figure 1 it is assumed that every individual has his or her individual specific reference point: Ref1 and Ref2. When selecting the best attribute level between $A_{1,3}$ and $A_{2,3}$ individual 1 and 2 will compare the utility value of each attribute level with the utility value of their individual specific reference point. Individual 1 is likely to select $A_{1,3}$ as best because of the larger utility differences:

$$\left(\text{Best} = A_{1,3}\right) = \left(\left(U_{1,3} - U_{\mathbf{Ref1},i=1}\right) > \left(U_{2,3} - U_{\mathbf{Ref2},i=1}\right)\right) \qquad \text{eq. 3}$$

At the same time individual 2 will be more likely to select $A_{2,3}$ as best because of the larger differences in utilities:

$$\left(\text{Best} = A_{2,3}\right) = \left(\left(U_{2,3} - U_{\mathbf{Ref2},i=2}\right) > \left(U_{1,3} - U_{\mathbf{Ref1},i=2}\right)\right) \qquad \text{eq. 4}$$

To avoid reference point differences affecting choices n BWS-2R, we can include explicit non-individual specific reference points in the choice task, which would lead to more precise preference measurements. In BWS-2R reference points across individuals are identical. This implies that for attribute 1 in Figure 2, respondents 1 and 2 will use the same reference point: $\text{Ref1}_{i=1} = \text{Ref1}_{i=2}$, and similarly for attribute 2. In the example in Figure 2, this will lead to the situation where both individuals will select $A_{1,3}$ over $A_{2,3}$ as best. Since the reference point is explicit and not individual specific, attribute level rankings will more accurately represent differences in preferences since they will no longer be confounded by differences in individual specific reference points.

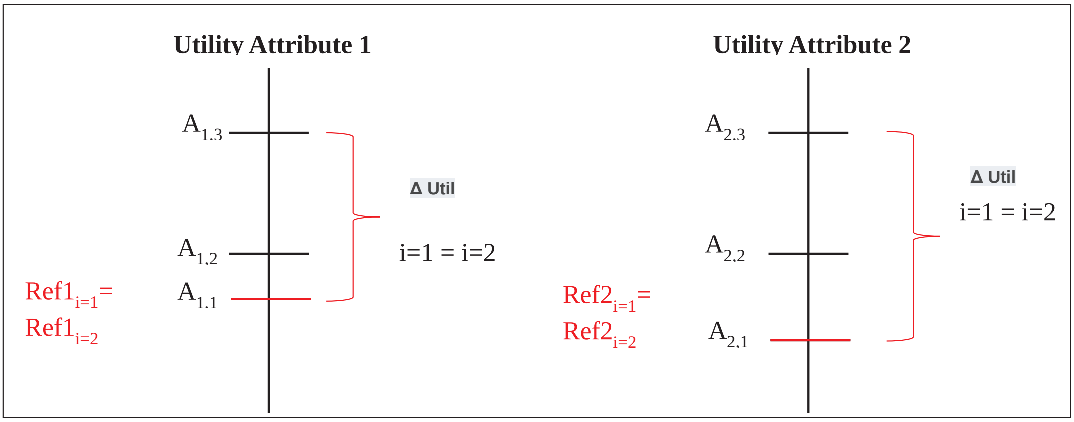


Figure 2 Utility levels of attribute levels with identical individual specific reference points

When including explicit reference points in BWS-2R choice tasks, the layout will change compared to traditional BWS-2 choice tasks (Figure 3a). In BWS-2R tasks it is much more clear what individuals receive when not selecting an attribute level as best, for example 40 percent effectiveness when focusing on this attribute. For BWS-2 tasks (Figure 3b) this is not clear from the choice task itself. Therefore BWS-2R provides individuals more information about the decision context, which makes the task itself less ambiguous.

Explicit reference point levels could overlap (reference point level equals attribute level) with attribute levels in the study design. In the example in Figure 1, Ref1 and Ref2 could be overlapping with the attribute levels mentioned in this example. For the sake of simplicity, we will focus on the situation when there is overlap and we set the reference level equal to the lowest attribute level utility (Figure 2). In the case of overlap, the attribute level that is used as reference point will automatically have a utility of 0, relative to the reference point:

Treatment characteristics	Changes	Most attractive	Least attractive
Effectiveness	From **40** to **60** percent	[ ]	[ ]
Number of pills a day	From **1** to **3** pills a day	[ ]	[ ]
Chance of getting an itchy skin	From **15** to **30** percent	[ ]	[ ]
Combination with alcohol	From **cannot** to **can be** combined	[ ]	[ ]

(a)

Treatment characteristics	Level	Most attractive	Least attractive
Effectiveness	60 percent	[ ]	[ ]
Number of pills a day	3 pills a day	[ ]	[ ]
Chance of getting an itchy skin	30 percent	[ ]	[ ]
Combination with alcohol	Can be combined	[ ]	[ ]

(b)

Figure 3 Example choice task with (a) and without (b) explicit reference points

$$\left(U_{k,lowest} - U_{Refk}\right) = \left(U_{Refk} - U_{Refk}\right) = 0 \qquad \text{eq. 5}$$

In this case, the attribute level utility (relative to the reference point) does not need to be inferred and no choice tasks including this attribute level are needed, leading to fewer choice tasks necessary in BWS-2R.

7.2.3 Interpretation of BWS-2R outcomes

Estimated parameters in BWS-2R can be interpreted in the same way as those in case 3 BWS or DCEs. This new form BWS-2 will be a less ambiguous choice experiment since the comparisons of attribute levels to reference points are explicitly defined and observed. Additionally, the parameter estimation procedure is generally the same as with BWS-2, where one reference category is needed for estimation. However, since each attribute level is considered in relation to its specific reference point, the interpretation of the BWS-2R outcomes will be different compared to (traditional) BWS-2. A full ranking of all attribute levels will not be possible because of the relationship between attribute level and specific reference points.

7.3 Empirical study design

7.3.1 Study population

We conducted an empirical study to investigate the impact of including reference points on BWS-2 outcomes. A study sample of the general population in the Netherlands from a Dutch Internet panel maintained by Right Minds B.V. was selected from June until July 2020. Since this study is based on a previous study from de Bekker-Grob et al.[18] about colorectal cancer screening (CRC screening), only individuals between 55 and 75 years (i.e., target population for CRC screening) were eligible to participate in this study. A minimum sample size of 300 for both BWS-2 as well as BWS-2R was regarded to be sufficiently large for reliable statistical analyses based on literature.[14,19]

7.3.2 Attributes and attribute levels

Attributes and levels for BWS-2R and BWS-2 were based on the previous study from de Bekker-Grob et al.[20], which provides more details about the selection procedures. For both BWS-2R and BWS-2 attributes included: effectiveness of the screening test, probability the screening test does not find the cancer (non-detection), waiting time for screening test result, waiting time for follow-up test and frequency of screening. Each attribute consisted of three levels as shown in Table 2.

Table 2 BWS-2 and BWS-2R attributes and attribute levels

Attributes	Levels
Effectiveness	20% 40% 60%
Probability screening test does not find the cancer	15% 25% 35%
Waiting time for test results	1 week 2 weeks 3 weeks
Waiting time for follow-up test	2 weeks 4 weeks 8 weeks
Frequency of screening	every year every 2 years every 3 years

7.3.3 Design of BWS choice tasks

For BWS-2R, an explicit reference point was included in the task. These reference points were set at the least attractive level for each attribute in Table 2. This means that every change from reference point to attribute level was positive, with more attractive attribute levels expected to being more preferred. In BWS-2, a regular BWS-2 approach was used without reference points in the choice tasks.

For both BWS-2 approaches an orthogonal main effect plan (OMEP) experimental design was used, which enables the independent estimation of preference weights for each attribute level.[6,14] For BWS-2, based on the number of attributes and levels, the minimum number of tasks needed to ensure no correlations between attributes was 16. We presented the 16 BWS-2 tasks to respondents and asked them to select their "most attractive" (i.e. best) and "least attractive" (i.e. worst) option in each task. Figure 4 illustrates an example BWS-2 task. For BWS-2R, a slightly different approach was used in developing the tasks. Since respondents might misinterpret or get frustrated with too many BWS-2R tasks with overlap between an attribute level and its reference point, the experimental design was adjusted in such a way that each BWS-2R task with overlap only included one attribute level and reference point that overlapped. To achieve this, choice tasks were generated using a smaller OMEP design excluding the reference points (i.e., two levels for each attribute), leading to an OMEP with 8 tasks which we mirrored for the next 8 tasks to get an equal number of 16 BWS-2R tasks compared to the 16 tasks in the BWS-2 study arm. For the last 8 tasks we five times randomly replaced an attribute level with the reference point to generate only five BWS-2R tasks in which attribute level and reference point overlap. An example BWS-2R task is shown in Figure 4.

CRC screening characteristics	Changes	Most attractive	Least attractive
Effectiveness	From **20** to **40** percent	☐	☐
Probability screening test does not find the cancer	From **35** to **25** percent	☐	☐
Waiting time for test results	From **3** to **1** week	☐	☐
Waiting time for follow-up test	From **8** to **4** weeks	☐	☐
Frequency of screening	From every **3** to every **2** years	☐	☐
Effectiveness	**40** percent	☐	☐
Probability screening test does not find the cancer	**25** percent	☐	☐
Waiting time for test results	**1** week	☐	☐
Waiting time for follow-up test	**4** weeks	☐	☐
Frequency of screening	**2** years	☐	☐

Figure 4 Example BWS-2R and BWS-2 choice tasks

7.3.4 Survey design

The survey consisted of eight sections: an introduction explaining the survey relevance and information about CRC screening, followed by background questions about sociodemographic characteristics and medical history, explanation of attributes and levels including questions for each attribute about respondent's self-reported reference points, part 1 of 8 BWS-2R/BWS-2 tasks, multiple questions about decision-making skills, part 2 of 8 BWS-2R/BWS-2 tasks, questions about preferences for BWS-2R or BWS-2 and final questions about the general mood of respondents and questions about the perceived difficulty of the survey. Respondents were randomly selected to answer to either BWS-2R or BWS-2. The survey was pilot tested with 34 respondents, using choice data and comments to open-ended feedback questions to optimize the survey. After the pilot, minor changes in the survey text were made. The survey was designed using Qualtrics XM survey software tool (Qualtrics, version June 2020).

7.3.5 Statistical analysis

Statistical analyses were conducted for each study arm of respondents who completed the survey. Given our interest in preference differences at an individual level in which explicit reference points in BWS-2R tasks will expectedly lead to less preference heterogeneity, we aimed to test if there is less preference heterogeneity when including reference points. Therefore, a mixed logit model (MXL) was employed for BWS-2 choice data analyses. Unlike the commonly used multinomial logit (MNL) model, MXL assumes a distribution of preference parameters across the sample which reflects differences in preferences among individuals.[17]

Using MXL, the utility (U) of an option in both BWS-2R and BWS-2 can be modeled as a linear function of the attributes and attribute levels, with

$$U = \sum_{k=1}^{A} \sum_{j=1}^{J_k} \beta_{i,k,j} X_{k,j} + \varepsilon \qquad \text{eq. 6}$$

where there are A attributes with attribute *k* having J_k attribute levels, with $X_{k,j}$ equal to one if the attribute level j of an attribute k is available in the presented BWS profile, $\beta_{i,k,j}$ is the individual *i* specific utility parameters for the j^{th} level of attribute *k*, and ε being the random error term representing the unexplained part of utility. The utility parameters for each attribute level are assumed to be normally distributed with mean $\beta_{j,k}$ and standard deviation $\sigma_{j,k}$.[21] For all BWS-2 estimations, scale differences between best and worst choices were allowed with the estimation of a best-worst scale parameter (beta_worst) in our model.[22,23]

Individuals' self-reported reference points were used to study whether these impacted choices. This was accomplished by allowing self-reported reference points to shift the attribute level specific utilities by subtracting them from each of the attribute level utilities (self-reported reference point adjusted MXL model). The utility for each attribute level is defined as:

$$U = (\beta_{i,k,j} - (\beta_k * SRR)) * X_{k,j} \qquad \text{eq. 7}$$

With $\beta_{i,k,j}$ the individual *i* specific utility parameters for the j^{th} level of attribute *k*, β_k the utility parameter for the attribute of interest, SRR the self-reported reference point value for the attribute of interest and $X_{k,j}$ equal to one if the attribute level j of an attribute k is available in the presented BWS profile.

All MXL models were programmed using R version 4.0.0 (Apollo package, version 0.0.1) using 1,000 Halton draws for estimation.[6,23] Dummy coding was implemented for all models, with 35% probability the screening test does not find the cancer as reference category (fixed at zero). Since in relative importance (RI) calculations scale differences between attributes cancel out, RI scores of attributes were also calculated and plotted for BWS-2R and BWS-2. RI scores were calculated by looking at the utility differences between the best and worst attribute level within a specific attribute. In order to gain insights into statistically significant differences in attribute RI scores between BWS-2R and BWS-2, we used a parametric bootstrap method to simulate 10,000 sets of model coefficients for all attribute levels from a multivariate normal distribution. Estimated coefficients and the variance-covariance matrix were used as inputs for the bootstrap. To test if BWS-2R and BWS-2 RI scores were statistically significantly different from each other, we tested whether in more than 97.5% or less than 2.5% of the 10,000 cases RI BWS-2R scores were larger than RI BWS-2 scores. If so, we concluded that there was evidence that BWS-2R and BWS-2 RI scores were statistically significant different from each other. Furthermore, time to complete and overall BWS-2R or BWS-2 preference analyses were conducted in order to compare the two approaches.

7.4 Empirical study results

In total 301 respondents completed the BWS-2R survey, while 307 respondents completed the BWS-2 survey. Finally, 2 (BWS-2R) and 5 (BWS-2) respondents were excluded for analyses since they selected attribute levels as both best and worst, leading to 299 (BWS-2R) and 302 (BWS-2) respondents included for final analyses. Sample characteristics for both study arms are summarized in Table 3, with both arms being overall comparable to each other except for relatively small differences in the number of females and age. Table 4 shows summary

Table 3 BWS-2R and BWS-2 sample characteristics including difference test p-values BWS-2R and BWS-2

Characteristic	Frequencies (%) BWS-2R (n=299)		BWS-2 (n=302)		Difference test p-value
Sex					0.02
Female	145	(49%)	174 (58%)		
Age (years)					0.01
Mean (min-max)	68.5 (55-75)		67.3 (55-75)		
Highest level of education					0.72
Elementary	3	(1%)	4	(1%)	
Secondary	168	(56%)	193	(64%)	
Higher Vocational	86	(29%)	70	(23%)	
University	29	(10%)	16	(6%)	
Other	13	(4%)	19	(6%)	
Health state score (score)*					0.64
Mean (min-max)	76.6 (0-100)		76.1 (10-100)		
General mood					0.87
Often nervous or restless	15	(6%)	20	(7%)	
Often disturbing thoughts	14	(5%)	18	(6%)	
Often able to let go worrying thoughts	102	(34%)	91	(30%)	
Often a relax feeling when thinking of recent concerns	106	(35%)	100	(33%)	
Hospital visits last month					0.47
Once	51	(17%)	60	(20%)	
Multiple times	26	(9%)	25	(8%)	
No	222	(74%)	217	(72%)	
History with cancer					0.81
Yes, colorectal cancer	9	(3%)	9	(3%)	
Yes, other type of cancer	43	(14%)	46	(15%)	
No	247	(83%)	247	(82%)	
History with screening programmes					0.51
Yes, colorectal cancer screening	174	(58%)	159	(53%)	
Yes, other screening programme	19	(7%)	20	(7%)	
Both CRC and other screening programme	79	(26%)	96	(31%)	
No	27	(9%)	27	(9%)	
CRC screening programme outcome					0.33
Postive outcome, no follow up	228	(76%)	237	(78%)	
Negative outcome, follow up needed	25	(8%)	18	(6%)	
Missing	46	(16%)	47	(16%)	
Health literacy (score) **					0.44
Mean comfortable with percentages (min-max)	4.1	(1-6)	4.1	(1-6)	
Mean comfortable with calculating tips (min-max)	3.9	(1-6)	3.9	(1-6)	

* Score 100 indicates best health state; score 0 indicates worst health state
**Mean score on a scale from 1-6, with 6 being the best score

Table 4 BWS-2R and BWS-2 self-reported reference point summary statistics including difference test p-values BWS-2R and BWS-2

Attribute	BWS-2R (n=299) Min	Max	Mean	SD	BWS-2 (n=302) Min	Max	Mean	SD	Difference test p-value
Effectiveness	2.44	100.00	47.59	26.42	2.44	100.00	49.05	26.92	0.56
Chance screening test does not find the cancer	0.00	100.00	20.22	20.73	0.00	90.00	21.36	20.32	0.37
Waiting time for test results	0.00	20.00	2.30	1.79	1.00	30.00	2.54	3.35	0.40
Waiting time for follow-up test	0.00	12.00	2.48	1.73	1.00	48.00	2.74	3.26	0.21
Frequency of screening	1.00	5.00	2.26	1.06	0.00	10.00	2.46	1.40	0.32

statistics of the self-reported reference point for both BWS-2R as well BWS-2. Results indicate no statistically significant differences between both study arms.

7.4.1 Mixed logit results

Table 5 presents the estimated preference weights (β's) and preference heterogeneity parameters (σ's) from the MXL estimation for BWS-2R and BWS-2, with the attribute level "35% probability the screening test does not find the cancer" as the reference level. For BWS-2R, the three attribute levels that were used as reference points in the BWS-2R choice tasks (i.e., 20% effectiveness, 8 weeks waiting time for follow-up test and screening every 3 years) were as expected not statistically significantly different from zero. Preference weights indicate respondents most preferred screening option with a 60% effectiveness, 15% chance the screening test does not find the cancer, with 1 week of waiting time for test results, 2 weeks of waiting time for follow-up test and screening every year. The BWS-2 results show a similar pattern, with 60% effectiveness being the most preferred attribute level, followed by 40% effectiveness and 1 week waiting time for test results. Focusing on the magnitudes of the estimated preference weights, more attractive attribute levels were preferred compared to less attractive levels for both BWS-2R and BWS-2.

To study if estimated preference weights differed between BWS-2R and BWS-2, they were tested to be statistically significant different from each other. Statistically significant differences (t-test, $p < 0.05$) were found for three attributes (60% effectiveness, 15% chance the screening test does not find the cancer and 25% chance the screening test does not find the cancer). Additionally, statistically significant differences were also found for two other attributes: 1 week waiting time for test results and 15% chance the screening test does not find the cancer (t-test, $p < 0.1$).

Since statistically significant differences between estimated BWS-2R and BWS-2 preference weights were found and in RI scale cancels out, RI scores were calculated using parametric bootstrap method and plotted in boxplots with red squares indicating the mean values (Figure 5). For BWS-2R effectiveness had the highest RI score, followed by waiting

Table 5 Mixed logit results for BWS-2R and BWS-2

	BWS-2R				BWS-2			
	β est.	σ est.	Rob. Se (β)	Rob. Se (σ)	β est.	σ est.	Rob. Se (β)	Rob. Se (σ)
Effectiveness (%)								
20[i]	0.33	1.24	0.19	0.66	2.49*	3.41*	0.40	0.31
40	2.90*	2.34*	0.27	0.21	4.30*	3.24*	0.42	0.29
60	4.44*	3.83*	0.36	0.58	5.98*	3.10*	0.39	0.53
Probability screening test does not find the cancer (%)								
15	2.11*	1.50*	0.25	0.21	1.55*	1.38*	0.22	0.20
25	1.33*	1.21*	0.17	0.18	0.47*	1.73*	0.14	0.16
35[i]	REF	0.01	-	0.25	REF	1.58*	-	0.27
Waiting time test results (weeks)								
1	1.64*	1.48*	0.18	0.19	3.77*	1.01*	0.30	0.13
2	0.96*	1.04*	0.16	0.16	2.72*	0.06	0.28	0.21
3[i]	0.38*	0.21	0.16	0.19	1.55*	0.25	0.21	0.24
Waiting time follow-up test (weeks)								
2	2.01*	0.62*	0.19	0.18	2.67*	0.30	0.28	0.16
4	1.12*	0.52*	0.16	0.20	1.11*	0.76*	0.20	0.15
8[i]	-0.29	0.12	0.15	0.16	0.08	0.41	0.16	0.40
Frequency of screening (every … years)								
1	1.91*	2.13*	0.25	0.18	3.06*	1.37*	0.32	0.14
2	1.25*	1.75*	0.23	0.23	2.26*	1.37*	0.25	0.11
3[i]	-0.28	0.17	0.16	0.30	0.72*	1.22*	0.23	0.22
beta_worst[ii]	1.18		0.10		0.99		0.08	
Sample size	299				302			
Log likelihood	-9577.48				-8891.52			

* significant at 5%
i attribute levels in BWS-2R used as explicit reference point
ii beta_worst parameter allows for scale differences between best and worst choices (hypothesis testing beta_worst = 1 showed no statistically significant outcomes)

time for follow-up test, frequency of screening, probability the screening test does not find the cancer and waiting time for test result. For BWS-2 effectiveness also had the highest RI score. This was however followed by waiting time for follow-up test, frequency of screening, waiting time for test result and probability the screening test does not find the cancer. Based on the bootstrapped RI scores testing procedure, BWS-2R RI scores were found to be statistically significantly different from BWS-2 RI scores for attributes: effectiveness and for probability the screening test does not find the cancer (higher BWS-2R RI score), as well as for waiting time for test result (lower BWS-2R RI score)

We expected less heterogeneity when explicit reference points were included in BWS choice tasks since in this case preference heterogeneity is entirely based on individuals' preferences and not confounded with heterogeneity in individuals' reference points.

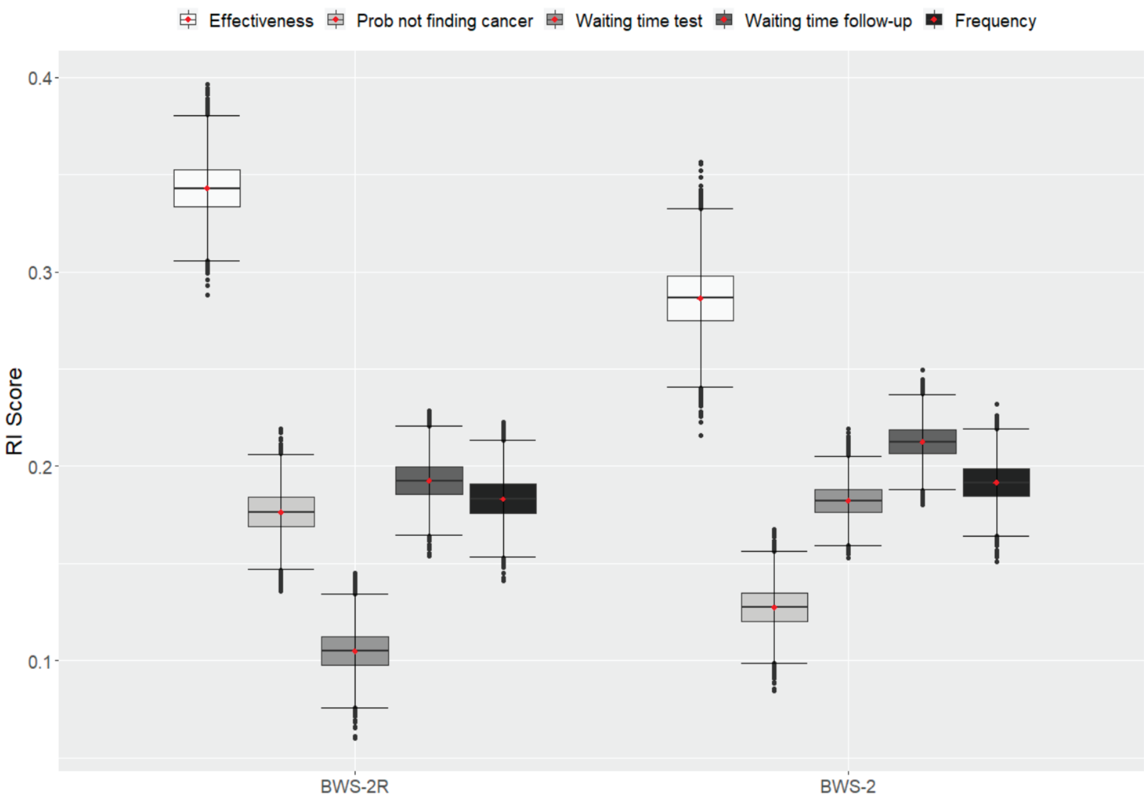


Figure 5 Relative importance score boxplots for BWS-2R and BWS-2. The red squares indicate mean scores.

Therefore, self-reported-reference-point-adjusted MXL models were estimated to see whether these self-reported reference points impacted the heterogeneity for both BWS-2R and BWS-2 compared the heterogeneity from the regular MXL models (Table 6). Results indicated that for the self-stated reference points in BWS-2R no attributes were statistically significant, meaning that there is no evidence that self-reported reference points did affect preferences in BWS-2R. This was expected since BWS-2R already included reference points in the choice task. In BWS-2 several attributes were statistically significant: non-detection attribute, waiting time test results and screening frequency. We did however not find strong differences in heterogeneity between the adjusted and regular MXL model outcomes, but the log-likelihood increased (as expected) in the adjusted model indicating a better model fit.

Table 6 Self-reported-reference-point-adjusted mixed logit heterogeneity estimates for BWS-2R and BWS-2

	BWS-2R self-reported reference point adjusted		BWS-2R		BWS-2 self-reported reference point adjusted		BWS-2	
	σ est.	Rob. Se (σ)	σ est.	Rob. Se (σ)	σ est.	Rob. Se (σ)	σ est.	Rob. Se (σ)
Effectiveness (%)								
20	1.00	0.78	1.24	0.66	3.27*	0.37	3.41*	0.31
40	2.66*	0.52	2.34*	0.21	3.42*	0.46	3.24*	0.29
60	3.13*	0.40	3.83*	0.58	2.77*	0.26	3.10*	0.53
Probability screening test does not find the cancer (%)[i]								
15	1.68*	0.20	1.50*	0.21	1.28*	0.18	1.38*	0.20
25	1.18*	0.31	1.21*	0.18	1.65*	0.26	1.73*	0.16
35	0.20	0.29	0.01	0.25	1.13*	0.34	1.58*	0.27
Waiting time test results (weeks)[i]								
1	1.60*	0.15	1.48*	0.19	0.94*	0.17	1.01*	0.13
2	1.04*	0.18	1.04*	0.16	0.29	0.30	0.06	0.21
3	0.15	0.21	0.21	0.19	0.02	0.61	0.25	0.24
Waiting time follow-up test (weeks)								
2	0.71*	0.18	0.62*	0.18	0.35	0.21	0.30	0.16
4	0.81*	0.31	0.52*	0.20	0.75*	0.15	0.76*	0.15
8	0.30	0.26	0.12	0.16	0.67*	0.19	0.41	0.40
Frequency of screening (every … years)[i]								
1	2.46*	0.36	2.13*	0.18	1.58*	0.18	1.37*	0.14
2	1.84*	0.28	1.75*	0.23	1.25*	0.12	1.37*	0.11
3	0.53	0.39	0.17	0.30	1.18*	0.26	1.22*	0.22
Sample size	299		299		302		302	
Log likelihood	-9516.70		-9577.48		-8878.80		-8891.52	

* significant at 5%
i attributes significant in this model specification for BWS-2

7.4.2 Survey completion time and preference

The average time to complete the BWS-2R survey was 23 minutes, while BWS-2 respondents completed the survey in 21 minutes on average. Wilcoxon rank-sum test results indicated that these averages were not statistically significant different from each other ($p > 0.05$). Figure 6 plots the average answering time for each BWS-2R and BWS-2 choice task. Differences between both BWS-2 approaches were most pronounced at the start and after the intermediate questions.

Table 7 shows the results for BWS-2R and BWS-2 regarding the overall difficulty of the survey and preferences for BWS-2R or BWS-2. For both groups most respondents evaluated the survey as neutral or easy and differences between the two BWS-2 approaches were not statistically significant (chi-squared test, $p > 0.05$). Asking respondents whether they preferred the BWS-2 approach they completed or the other approach, a larger proportion of respondents that completed BWS-2 preferred BWS-2R (48%) compared to the proportion that completed BWS-2R and preferred BWS-2 (41%). Chi-squared test results showed that these differences are statistically significant ($p < 0.05$).

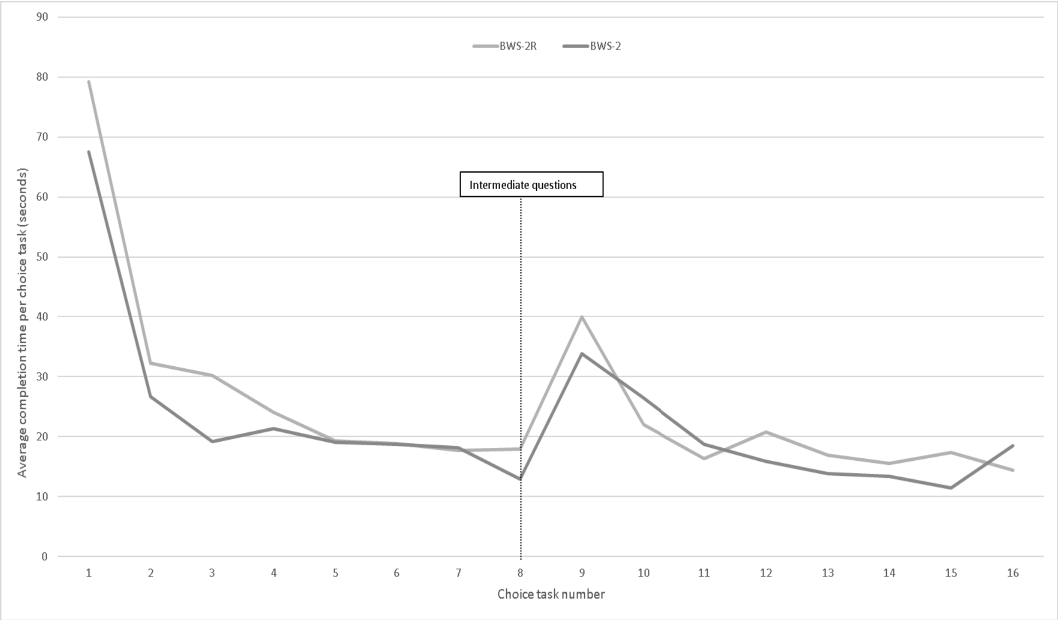


Figure 6 Average completion time per choice task for BWS-2R and BWS-2

Table 7 Difficulty of the survey and preferences for BWS approach including difference test p-values

Characteristic	Percentages: BWS-2R (n=299)	BWS-2 (n=302)	Difference test p-value
Difficulty survey			0.47
Very easy	5%	6%	
Easy	19%	23%	
Neutral	56%	54%	
Difficult	18%	16%	
Very difficult	3%	1%	
BWS approach preference			0.01
BWS-2R	59%	48%	
BWS-2	41%	52%	

7.5 Discussion

In this study we investigated the benefits of BWS-2R, which includes explicit reference points in BWS-2 to overcome results that are driven by (unobserved) differences in reference points. We conclude that by using BWS-2R instead of BWS-2, interpretation of outcomes will be clearer and also reduces noise in the inferred preferences as regular BWS-2 preferences are confounded with the reference points. In both BWS-2R and BWS-2 effectiveness was the most important attribute. Statistically significant differences between estimated preference weights for both BWS-2 approaches were found. Looking at the self-stated-reference-point-adjusted estimation outcomes for BWS-2 the self-reported reference point adjusted model indicated a better model fit compared to the regular model, although we did not find strong evidence for less heterogeneity compared at the regular model (not adjusted for self-reported reference points). We did however find evidence for statistically significant differences in RI scores between BWS-2R and BWS-2 for attributes effectiveness, probability the screening test does not find the cancer and waiting time for test result. (α= 0.05). Analysis also showed no difference in perceived difficulty between BWS-2R and BWS-2, with a larger proportion of respondents that completed BWS-2 preferred BWS-2R compared to the proportion that completed BWS-2R and preferred BWS-2.

One of the main findings of this study was that introducing BWS-2R, i.e. BWS-2, with explicit reference points, leads to results that are not sensitive to (differences in) individuals' reference points. Our empirical results showed statistically significant differences in estimated preference weights between BWS-2R and BWS-2, for example for the effectiveness and non-detection attributes. This is in line with our theoretical expectations. Specifically, we expected that respondents would find it more difficult to imagine a reference point for the non-detection attribute, as this requires more in-depth knowledge of CRC screening, compared to, for example, the waiting time for test results. When the reference point is not so easy to imagine and hence the respondent does not know what s/he gets when not selecting that attribute, it is expected that the specific attribute will have less impact. This is exactly what the RI plot shows: the non-detection attribute is ranked lower in BWS-2 compared to BWS-2R. Using explicit reference points in BWS-2R alleviates the biases in attribute importance that could arise from reference point ambiguity.

Based on theory we expected that BWS-2R leads to a clear interpretation of the estimation results, unaffected by heterogeneity in or ambiguity of the reference points that individuals have. From the perspective of the analyst, BWS-2R will therefore be preferred, since the decision-making process is less of a black box: the analyst knows the point (explicit reference point) the attribute levels are compared to when individuals make their best and worst choices. Preference heterogeneity parameters (σ's) for BWS-2R do not capture reference point heterogeneity since these reference points are fixed. These parameters therefore more accurately capture true heterogeneity in preferences in the case of BWS-2R. This is in line with theoretical expectations and in line with a DCE preference study from Mao et al.[24], which showed that reference-dependent models are more appropriate to model individuals' choices, especially when aiming to explain the origin of preference heterogeneity. DCE studies by for example Stathopoulos & Hess[25] and Hess et al.[26] also describe the impact of reference points on preferences of individuals.

A strength of this study is that it is the first study that introduces BWS-2 with explicit reference point leading to a more accurate representation of preferences, therefore improving existing (health) preference elicitation methods. We believe this is especially important in the context of BWS-2 since individuals make choices at the attribute level instead of the profile level. As a result, no trade-offs are made across differences in attribute values, like for example in DCE or BWS-3, where the reference points would cancel out in the evaluation.

A potential limitation of this study is that in BWS-2R (compared to BWS-2) a full ranking of all attribute levels is not possible due to the relationship between attribute level and specific reference point, with attribute levels no longer being directly comparable. However, differences between attribute levels can still be interpreted and we believe the problems associated with completely unknown reference points in BWS-2 will be much larger than the DCE-like interpretation BWS-2R demands.

To conclude, in this study it was shown that including explicit reference points in BWS-2 improves outcomes as it clarifies interpretation of the estimation results. It also resulted in a different and we argue more accurate ranking of RI scores as regular BWS-2 rankings are confounded with reference points. This study also showed no difference in perceived difficulty between BWS-2R and BWS-2, with respondents preferring BWS-2R over BWS-2. Hence, we advise using BWS-2R when aiming to conduct a BWS-2 experiment for preference research in health economics.

References

1. Bridges JFP, 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 Heal*. 2011;14(4):403-413. doi:10.1016/j.jval.2010.11.013
2. Food and Drug Administration (FDA). *Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling*.; 2016. Accessed July 25, 2018. http://www.regulations.gov.
3. MDIC. MEDICAL DEVICE INNOVATION CONSORTIUM PATIENT CENTERED BENEFIT-RISK PROJECT REPORT: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. *Med Device Innov Consort*. Published online 2015. https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiKk5fwtIDSAhUCvRoKHaC_CoIQFgggMAA&url=http%3A%2F%2Fwww.fda.gov%2Fdownloads%2FScienceResearch%2FSpecialTopics%2FRegulatoryScience%2FUCM486253.pdf&usg=AFQjCNFbkufZYkBrW
4. Mahieu P-A, Andersson H, Beaumais O, Crastes dit Sourd R, Hess S, Wolff F-C. Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. *Rev Agric Food Environ Stud*. 2017;98(3):201-220. doi:10.1007/s41130-017-0053-6
5. Soekhai V, Whichello C, Levitan B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324-1331. doi:10.1016/j.drudis.2019.05.001
6. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-189. doi:10.1016/j.jhealeco.2006.04.002
7. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):1-14. doi:10.1186/s13561-015-0079-x
8. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*.; 2015. doi:10.1017/CBO9781107337855
9. Thurstone LL. A law of comparative judgment. *Psychol Rev*. 1927;34(4):273-286. doi:10.1037/h0070288
10. 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. *Med Decis Mak*. 2014;34(5):638-654. doi:10.1177/0272989X14526640
11. van Dijk JD, Groothuis-Oudshoorn CGM, Marshall DA, IJzerman MJ. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders’ Risk Tolerance for Hip Replacement Surgery. *Value Heal*. 2016;19(4):316-322. doi:10.1016/j.jval.2015.12.020
12. Soekhai V, Donkers B, de Bekker-Grob E. PNS295 BEST WORST SCALING: FOR GOOD OR FOR BAD BUT NOT FOR BOTH. *Value Heal*. 2019;22:S813. doi:10.1016/j.jval.2019.09.2195
13. Werner KM, Zank H. A revealed reference point for prospect theory. *Econ Theory*. 2019;67(4):731-773. doi:10.1007/s00199-017-1096-2
14. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press; 2015. doi:10.1017/CBO9781107337855
15. Ben-Akiva ME, Lerman SR. *Discrete Choice Analysis : Theory and Application to Travel Demand*. MIT Press; 1985.
16. McFadden D. *Conditional Logit Analysis of Qualitative Choice Behaviour*. Vol 1. (Zarembka P, ed.). Academic Press; 1973. doi:10.1108/eb028592
17. Mcfadden D. *Economic Choices*. Vol 91.; 2001. Accessed April 9, 2019. http://users.auth.gr/cemman/teaching/QuantMeth/McFadden2001.pdf
18. de Bekker-Grob EW, Donkers B, Veldwijk J, et al. What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. *Patient*. 2020;(0123456789). doi:10.1007/s40271-020-00477-w
19. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *Patient*. 2015;8(5):373-384. doi:10.1007/s40271-015-0118-z
20. de Bekker-Grob EW, Donkers B, Veldwijk J, et al. What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. *Patient*. 2020;(0123456789). doi:10.1007/s40271-020-00477-w
21. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical Methods for the Analysis of

Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Heal*. 2016;19(4):300-315. doi:10.1016/j.jval.2016.04.004
22. Hess S, Palma D. Apollo : a flexible , powerful and customisable freeware package for choice model estimation and application Stephane Hess & David Palma Choice Modelling Centre University of Leeds. *Apollo user Man*. 2019;(April).
23. Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *J Choice Model*. 2019;32:1-43. doi:10.1016/j.jocm.2019.100170
24. Mao B, Ao C, Wang J, Xu L. The Importance of Loss Aversion in Public Preferences for Wetland Management Policies: Evidence from a Choice Experiment with Reference-Dependent Discrete Choice Model. Wetlands. 2020;40(3):599-608. doi:10.1007/s13157-019-01195-2
25. Stathopoulos A, Hess S. Revisiting reference point formation, gains-losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. Transp Res Part A Policy Pract. 2012;46(10):1673-1689. doi:10.1016/j.tra.2012.08.005
26. Hess S, Rose JM, Hensher DA. Asymmetric preference formation in willingness to pay estimates in discrete choice models. Transp Res Part E Logist Transp Rev. 2008;44(5):847-863. doi:10.1016/j.tre.2007.06.002

Appendix A

a

Best		Worst
[ ]	Being cured	[ ]
[ ]	Severe side effects	[ ]
[ ]	Voice changes	[ ]
[ ]	Calcium deficiency	[ ]

b

Best		Worst
[ ]	Chance of being cured: 40%	[ ]
[ ]	Chance of severe side effects: 2%	[ ]
[ ]	Chance of voice changes: 5%	[ ]
[ ]	Chance of calcium deficiency: 5%	[ ]

c

Treatment A	Treatment B	Treatment C	
Chance of being cured: 40%	Chance of being cured: 70%	Chance of being cured: 100%	
Chance of severe side effects: 2%	Chance of severe side effects: 10%	Chance of severe side effects: 5%	
Chance of voice changes: 5%	Chance of voice changes: 0%	Chance of voice changes: 10%	
Chance of calcium deficiency: 5%	Chance of calcium deficiency: 10%	Chance of calcium deficiency: 15%	
[ ]	[ ]	[ ]	Best
[ ]	[ ]	[ ]	Worst

Figure A Examples of the three BWS cases, with a. case 1 BWS (object case), b. case 2 BWS (profile case) and c. case 3 BWS (multi-profile case)

Part III

Empirically comparing outcomes between DCE and BWS-2

Chapter 8

What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting

Himmler S, Soekhai V, van Exel J and Brouwer W

Published Journal of Choice Modelling 2021

Abstract

To appropriately weight dimensions of quality-of-life instruments for health economic evaluations, population and patient preferences need to be elicited. Two commonly used elicitation methods for this purpose are discrete choice experiments (DCE) and case 2 best-worst scaling (BWS). These methods differ in terms of their cognitive burden, which is especially relevant when eliciting preferences among older people. Using a randomized experiment with respondents from an online panel, this paper examines the cognitive burden associated with color-coded and level overlapped DCE, color-coded BWS, and 'standard' BWS choice tasks in a complex health state valuation setting. Our sample included 469 individuals aged 65 and above. Based on both revealed and stated cognitive burden, we found that the DCE tasks were less cognitively burdensome than case 2 BWS. Color coding case 2 BWS cannot be recommended as its effect on cognitive burden was less clear and the color coding lead to undesired choice heuristics. Our results have implications for future health state valuations of complex quality of life instruments and at least serve as an example of assessing cognitive burden associated with different types of choice experiments.

8.1 Introduction

Developments like ageing populations and rapid advances in medical technology create challenges for budgets of publicly funded health care systems (de Meijer et al, 2013). Policy makers increasingly have to decide about which health care services to include in the basic benefits package, which should only be made available to certain subpopulations, and which should not be funded at all. Health technology assessment (HTA) generates valuable insights to support this decision-making process, using tools like cost-utility analysis. There, the benefits of health technologies are typically expressed in the incremental amount of health changes they produce. This is calculated based on data from generic, multidimensional quality of life instruments, and a weighting algorithm for the levels of the dimensions based on population or patient preferences (Neumann et al., 2016). Given that health and social care, for instance aimed at older persons, may affect more than health-related quality of life alone, more recently, broader well-being measure have been developed (Makai et al., 2014). These could facilitate cost-utility analyses with a broader scope in terms of relevant outcomes but require obtaining preferences for different 'well-being states' ideally anchored on death.

The measurement of population and patient preferences in health care is a rapidly developing field, with a plethora of qualitative and quantitative methods to the disposal of researchers and practitioners (Soekhai et al., 2019). One of the most popular methods over the last decade was the discrete choice experiment (DCE). Increasingly, population and patient preferences in health care are obtained using DCEs (Soekhai et al., 2018). The 'standard' DCE entails asking respondents to choose between two or more alternatives (Ryan et al., 2008) and is widely used for weighting quality of life instruments (Mulhern et al., 2018).[x] Another preference elicitation approach that gained traction over the last years also in this context, is best-worst scaling (BWS). There are three different forms of BWS – object case,

[x] This is also known as 'health state valuation'.

profile case, and multi-profile case. The following will focus on profile case, or also called case 2 BWS, where individuals have to select a best and a worst option from a list of dimension levels or items (Flynn and Marley, 2014). Case 2 BWS was applied to value different quality of life instruments before (Cheung et al., 2016). This includes the ICECAP-O, a well-being measure specifically aimed at older people (Coast et al., 2008).

While both DCE and BWS provide numerical estimates of the relative importance of the different levels and dimensions of the respective quality of life or well-being instrument, previous research directly comparing DCE and BWS has shown that the choice between these approaches is not neutral as resulting preference estimates can differ (see e.g. Krucien et al., 2017). According to a recent review comparing DCE and BWS, there seems to be no conclusive evidence yet on which of the methods should be preferred in terms of the validity of the estimates (Whitty and Oliveira Gonçalves, 2018). Both methods assume different choice processes and ultimately may be seen to answer more or less subtly different questions. Some researchers prefer DCEs because the modelled choice processes have a strong theoretical foundation in random utility theory (Louviere, 2004). Providing choices between multiple alternative profiles can also be considered as a more realistic way of the decision-making process compared to selecting a best and worst option from a list of items. Another advantage of DCEs in the context of health state valuation, is that utilities can more easily anchored onto the full health (or well-being)-dead scale. On the other hand, some argue that profile case BWS is to be preferred as it is a more efficient way of collecting data compared to DCE since each task entails two choices. Moreover, cognitive burden of BWS tasks may be lower, since individuals only need to focus on one set of attributes and levels in each choice task, compared to multiple in DCEs. Some specifically claim that it would be recommendable to choose case 2 BWS if DCE tasks are considered to be too burdensome (Flynn, 2010; Potoglou et al., 2011). However, Whitty and Oliveira Gonçalves (2018) conclude that there is no clear evidence for an advantage of BWS regarding participant acceptability in terms of feasibility of administration or response efficiency. The response efficiency, that is, the cognitive burden associated with choice tasks, is important as it influences choice consistency, respondent fatigue and the use of simplifying choice heuristics (Jonker et al., 2019), which could subsequently influence the validity of the preference estimates.

Due to the ageing of the population, the need for economic evaluations of health and social care services targeted at older people can be expected to increase. This makes accurately measuring and weighting quality of life dimensions in this population very important, and choosing the appropriate methodology to do so, all the more relevant. If one decides, as we do here, that an instrument aimed at older people should be weighted using older peoples' preferences,[xi] one needs to be aware of an additional aspect: Since there is a large variation in the level of cognitive abilities within older people, the design of choice experiments for this population should especially be wary of the complexity and subsequent cognitive burden of the choice task format in order to enable obtaining valid and reliable responses (Milte et al., 2014). Measuring and weighting quality of life or well-being outcomes inaccurately may ultimately lead to sub-optimal policy recommendations for resource allocation to health or social care services aimed at older people.

Specific evidence about the cognitive burden of DCE and case 2 BWS for valuing quality of life measures among older people is lacking. Therefore, the main aim of this study was to assess the cognitive burden and incidence of simplifying choice heuristics in DCE and case 2 BWS choice tasks among older people in this context. Another aim was to test the impact of the use of color coding on the cognitive burden and choice behavior of case 2 BWS tasks, which has been assessed for DCEs before (Jonker et al., 2019).

[xi] Whose values to elicit is debatable in health state valuation in general. We decided to use older peoples' preferences as the WOOP is intended to inform allocation decisions only within care for older people.

8.2 Methods

We set up a randomized experiment with three study arms to examine the cognitive burden and choice behavior attached to three respective choice task formats for valuing a quality-of-life instrument: a color coded and level overlapped DCE (5 out of 9 dimensions), a case 2 BWS and a color coded case 2 BWS.[xii] In the applied color coding, five shades of one color correspond to the five levels of attributes of the used instrument, with darker shades representing the least desirable levels. The rationale behind this type of coding in the DCE is that it helps respondents to identify differences between the alternatives, and higher and lower levels, while not nudging respondents to only focus on the differing attributes, what e.g. exclusively highlighting the non-overlapped levels would do, or introducing strong prejudgments on the severity of the levels by using e.g. a traffic light color coding.

We chose an online setting with participants from an online panel for our study, as this administration and sampling mode facilitates reaching a sufficiently large number of respondents for health state valuation studies, which is also why it is used in most such studies by now (Mulhern et al., 2018).

The quality-of-life measure used in the experiment was the recently developed Well-being of Older People instrument (WOOP) (Hackert et al., 2019). Examining the cognitive burden of a valuation task is especially important in the context of this new instrument for measuring the general/overall quality of life of older people: First, the WOOP consists of nine dimensions with five levels each, which requires complex choice tasks. Second, as preferences should be based on an older population, cognitive burden is of special relevance. The profiles shown to respondents in both DCE and BWS tasks corresponded to well-being states, described using the nine dimensions of the WOOP (i.e. physical health, mental health, social life, receive support, acceptance and resilience, feeling useful, independence, making

[xii] In the remainder of the paper, BWS refers to case 2 BWS.

ends meet, living situation).[xiii] In designing the choice tasks and their visual representation, we followed methodological work on the use of color coding and level overlap in DCEs aimed to reduce task complexity (Jonker et al., 2019, 2018; Maddala et al., 2003). To enable a more direct comparison and to test the impact of color coding on task complexity in BWS, which has not been studied before, the randomized experiment included a color coded BWS and a regular BWS.

Important to note here is that the design was generated to test the cognitive burden and choice behavior of older people, not to provide model estimates for the different methods. Due to the large descriptive system of the WOOP, this would have required estimation of 36 parameters in the DCE and 45 parameters in the BWS, a blocked design and a much larger sample size. While a comparison of model estimates would have been interesting, this was not our current research aim.

8.2.1. Survey structure and randomisation

The structure of the experimental survey is shown in Fig. 1. First, respondents were asked to complete the WOOP instrument to become familiar with its dimensions and levels. Afterwards, they were randomized 1:1:1 to the three study arms: color coded DCE (1), color coded BWS or BWSc (2), and regular BWS (3). The randomization was preferred over having the same respondents completing both DCE and BWS tasks, to have avoid the different parts of the experiment influencing each other and to stay as close as possible to standard DCE and BWS experiments. Furthermore, two full sets of valuation tasks per respondents were considered to be too burdensome. Respondents were familiarized with the presentation of well-being states in the subsequent experiment by showing them their own profile in DCE or BWS format based on the answers they previously gave to the WOOP

[xiii] Appendix A contains an updated version of the full descriptive system of the WOOP, with some formulations differing slightly compared to the version used in this study.

instrument. The choice task formats were introduced by a simple DCE or BWS task, where participants had to select between two types of fruits or chose the best and worst type of fruit from a list. The second part of the warm-up comprised of a choice task, as used in the subsequent experiment, providing further instructions. Subsequently, a block of six choice tasks was administered, followed by two simple break questions on an unrelated topic to interrupt the monotony and reduce respondent fatigue of answering the choice tasks. Then, a second block containing seven tasks concluded the randomized part of the questionnaire, leading to a total of 13 choice tasks per respondent. All respondents subsequently had to fill in three blocks of evaluation questions on a 5-point Likert scale, before providing some sociodemographic information at the end of the survey.

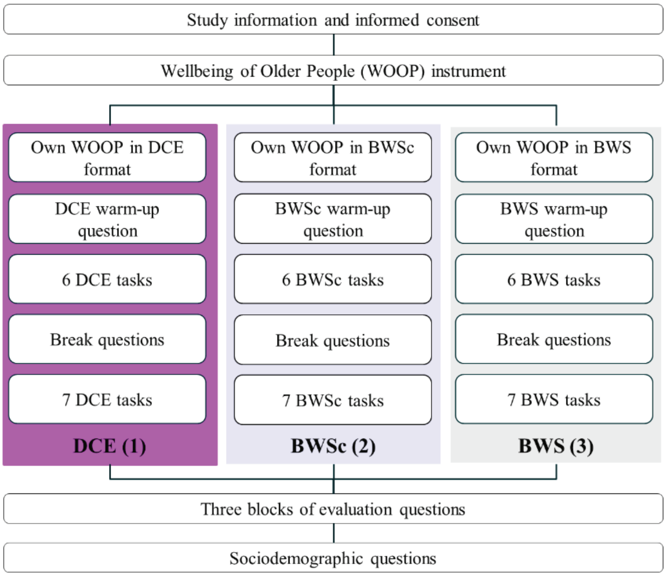


Fig. 1. Survey sturcutre and experimental arms

8.2.2. Survey administration and participants

The survey was programmed using Sawtooth software version 9.7.2 (Sequim, WA). We used Prolific.co to recruit survey participants, a platform for online subject recruitment specifically for research purposes (Palan and Schitter, 2018). Given our aim to assess the cognitive burden of the choice tasks in a sample of older people, being aged 65 or above was used as inclusion criteria (which is also the target population of the WOOP). Since this age group was underrepresented in the online panel, we had to combine respondents from the two largest country panels of Prolific.co, UK and U.S. residents, to obtain a reasonably sized sample. At the time of data collection, in October 2019, the potential respondent pool contained around 1,000 individuals. Using quota sampling, we aimed for 150 respondents for each of the three study arms. Respondents received a monetary compensation for participating, which was oriented on the mean completion time and averaged to an aggregated hourly reward of £7.62. To test the functionality of the survey and whether respondents understood the choice tasks, six think-aloud interviews with UK residents aged 65 and above were conducted (two per study arm) prior to the main data collection. These interviews showed that participants understood and appropriately engaged in the choice tasks (i.e. traded-off or considered multiple items).

8.2.3. Experimental design of DCE and BWS

Attributes and levels in the DCE and items of the BWS were based on the dimensions and levels of the WOOP instrument (Appendix A). This created a rather complex DCE setup with nine dimensions with five levels each and a BWS instrument with 45 items. WOOP well-being states were consequently defined by selecting one of the five levels from each of the nine dimensions for both DCE and BWS. In the DCE, respondents were repeatedly presented with two well-being states and asked to indicate, which of the two they preferred. An opt-out option was not included as this is uncommon in DCEs for health state valuation (Mulhern et al., 2018). In the BWS, a list of nine well-being items corresponding to one well-being state was shown to respondents. Participants then had to select the aspect that they most preferred (best) and the aspect that they least preferred (worst). ‘Most’ and ‘least’ is one of the options that are used for describing a best and worst choice (Huynh et al., 2017).[xiv]

To ensure that the choice tasks had a similar level of complexity compared to a regular choice experiment, choice tasks were created using standard design methodology as outlined in the subsequent paragraph. The literature on health related DCEs specifically targeted at older people was reviewed (in total 22 papers were studied) to inform the number of choice tasks. The number of choice tasks per respondent varied between 6 and 16 with a mean of 9.2. We opted to select a number of choice tasks at the upper end of this range (13) to capture fatigue effects (examples of this literature are e.g. Arendts et al., 2017; Franco et al., 2016; Milte et al., 2014) and because we anticipated this might be close to the approximate number in the actual valuation study of the WOOP. The 13 choice tasks consisted of 10 DCE choice tasks, two that repeat one of them, and one choice task to test for dominance.

The ten DCE choice were selected with help of Ngene design software (Version 1.2.1). To accommodate for level overlap (five out of the nine dimensions), which has been shown to reduce task complexity by Maddala et al. (2003) and Jonker et al. (2018), Ngene required a dataset including all possible candidate sets, i.e. combinations of two health states with five overlapped levels. To pragmatically reduce this to a feasible number, 5,000 out of the 1,953,125 possible health states were randomly selected and combined in MATLAB (MathWorks). Out of the obtained 25 million possible sets, we excluded the ones without the specified amount of overlap and randomly selected 1,000 sets out of the remaining 386,030 overlapped sets. Ngene was then used to select 10 choice tasks out of the 1,000 candidate sets

[xiv] ‘Most’ and ‘least’ may have a slightly different interpretation than ‘best’ and ‘worst’, but this should not have an impact on cognitive burden.

by optimizing for a conditional logit, main effects model (Appendix C contains the utility function) with 36 parameters corresponding to four of the five levels of each of the nine dimensions of the WOOP instrument. Small priors ranging from 0 to –0.25 were assumed, following the logical ordering of the WOOP levels. Besides the think-aloud interviews no further pilot testing was conducted.

An orthogonal main effects plan using Sawtooth software version 9.7.2 (Sequim, WA) was applied to generate 1,000 blocks of 10 choice tasks for the BWS experiment. Multiple levels from the same WOOP dimension were prohibited to appear in the same task. Following Flynn et al. (2015), to prevent uninformative sets, we reduced the occurrences of tasks with either only one top or bottom WOOP level by deleting all versions where this occurred more than 3 times in the 10 tasks. Out of the remaining 78 versions, one version was randomly selected to be used in the experiment.

We selected one of the created DCE and BWS choice tasks to appear as the second choice task and repeated the tasks at position 8 and 13, to test choice consistency, adding two choice tasks to the original 10 created tasks. In order to reduce the amount of noise in the answers, we chose tasks, which were expected to have a certain degree of utility difference between profiles in the DCE arm or provided somewhat clear BWS choices (the repeated choice tasks are shown in Appendix B). When this task was repeated the second time, the intensity color coding of the BWS task was intentionally reversed, to mislead respondents in order to assess the dependence on the color codes. A dominant DCE choice task and a BWS task, which was expected to have a clear best and worst choice were additionally created and added at position 6 to test the attention level of respondents, adding a third and final choice task to the original ten created tasks.[xv] The order of the dimensions (or attributes) was the same for all respondents within elicitation method and fixed for both DCE and BWS tasks to further

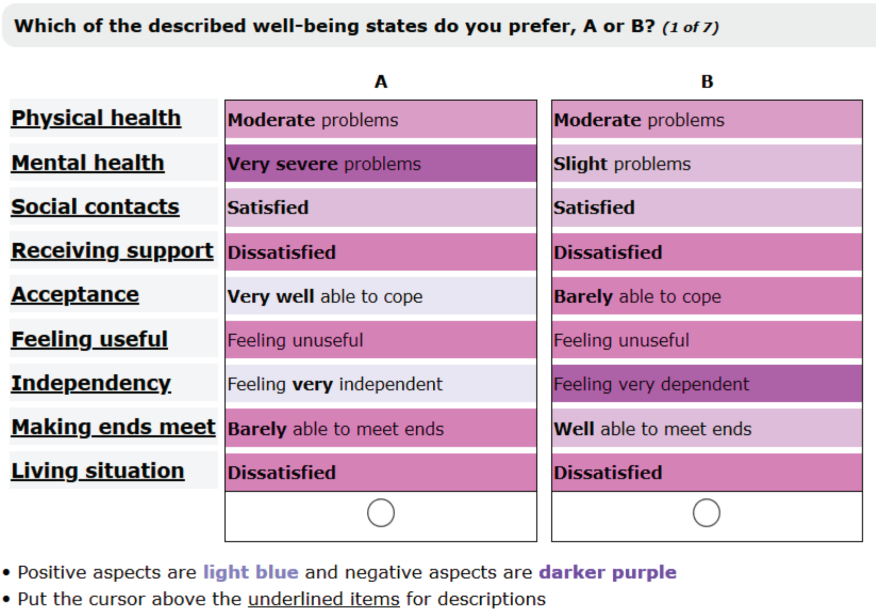
[xv] We decided against including results of this task in the final analysis, as such tasks are inherently difficult to compare between DCE and BWS (Whitty et al., 2018).

reduce task complexity. The only difference in attribute order between DCE and BWS tasks was that physical and mental health attributes where positioned in the middle of the BWS tasks, as we anticipated that these would be important dimensions and wanted to avoid respondents making their best and worst choice merely on the top without going over the remaining items. All respondents received the same 13 DCE tasks in study arm 1. Respondents in study arms 2 and 3 received the same 13 BWS tasks.

8.2.4 Visual presentation of choice tasks

The general visual representation of the choice tasks followed current practice, with the exception that intensity color coding was added to the choice tasks in study arms 1 and 2. Different shades of purple represented the different attribute levels, with the darker shades of purple highlighting the worse and the lighter shades and light blue expressing the better WOOP attribute levels in both the DCE and the color coded BWS tasks. In the explanation of the color coding in the survey, ‘better levels’ (e.g. very well able to cope, feeling very independent, no problems with physical health) were formulated as ‘positive aspects’ and ‘worse levels’ (e.g. barely able to cope, feeling very dependent, severe problems with physical health) as ‘negative aspects’ (e.g. Fig. 2). This type of color coding was previously used for DCEs by Jonker et al. (2017, 2018, 2019) and was found to reduce task complexity as well as attribute non-attendance, and was especially effective in combination with attribute level overlap. It was also shown that color-coding does not introduce bias in the choices and does not affect the relative importance of attributes (Jonker et al., 2019). The purple color scheme was specifically designed to accommodate for the most prevalent forms of color blindness. Additionally, shades of purple do not prompt natural or perceived value judgements, as opposed to for example traffic light color coding.

Fig. 2 shows an example of the layout of the color-coded (light blue to deep purple) and overlapped (five out of the nine dimensions) DCE choice task. Level descriptions of the WOOP instrument (Appendix A) were shortened for clarity, level labels were highlighted in bold, and attribute descriptions appeared merely as mouseovers on the attribute labels to reduce the amount of text. Fig. 3 shows examples of both color coded and non-color coded BWS tasks. Descriptions of attributes were also included as mouseovers, while the item text contained the full WOOP level descriptions.



Which of the described well-being states do you prefer, A or B? *(1 of 7)*

	A	B
Physical health	**Moderate** problems	**Moderate** problems
Mental health	**Very severe** problems	**Slight** problems
Social contacts	**Satisfied**	**Satisfied**
Receiving support	**Dissatisfied**	**Dissatisfied**
Acceptance	**Very well** able to cope	**Barely** able to cope
Feeling useful	Feeling unuseful	Feeling unuseful
Independency	Feeling **very** independent	Feeling very dependent
Making ends meet	**Barely** able to meet ends	**Well** able to meet ends
Living situation	**Dissatisfied**	**Dissatisfied**
	○	○

- Positive aspects are **light blue** and negative aspects are **darker purple**
- Put the cursor above the underlined items for descriptions

Fig. 2. Visual presentation of DCE choice task with color coding and level overlap.

Imagine living in this well-being state and select which aspect you would most prefer, and which aspect you would least prefer. ***(1 of 6)***

Most	Well-being state	Least
◯	I am dissatisfied with my social contacts	◯
◯	I am reasonably satisfied with the support I receive	◯
◯	I am reasonable able to deal with my circumstances and changes therein	◯
◯	I feel reasonably useful	◯
◯	I have slight problems with my physical health	◯
◯	I have very severe problems with my mental health	◯
◯	I feel very dependent	◯
◯	I am very well able to make ends meet	◯
◯	I am dissatisfied with my living situation	◯

- Positive aspects are **light blue** and negative aspects are **darker purple**
- Put the cursor above the underlined items for descriptions

Imagine living in this well-being state and select which aspect you would most prefer, and which aspect you would least prefer. ***(1 of 6)***

Most	Well-being state	Least
◯	I am dissatisfied with my social contacts	◯
◯	I am reasonably satisfied with the support I receive	◯
◯	I am reasonable able to deal with my circumstances and changes therein	◯
◯	I feel reasonably useful	◯
◯	I have slight problems with my physical health	◯
◯	I have very severe problems with my mental health	◯
◯	I feel very dependent	◯
◯	I am very well able to make ends meet	◯
◯	I am dissatisfied with my living situation	◯

- Put the cursor above the underlined items for descriptions

Fig. 3. Visual presentation of color-coded and non-color-coded BWS choice task.

8.2.5 Statistical analysis

To assess and compare the cognitive burden and possible choice heuristics associated with the three formats of choice tasks, three types of data were analyzed. First, objective measures including mean choice task completion time, development of time per task (assessing learning effects) and drop-out rates were calculated and compared. Second, mean response scores of the three blocks of debriefing questions on perceived choice complexity, the number of choice tasks, and choice strategies used, were obtained. The latter aimed to identify the extent to which respondents engaged in simplifying choice heuristics. This included two statements relating to the number of attributes commonly considered during the choice tasks, also known as attribute non-attendance (Yao et al., 2015), and a statement on deciding that all attributes/dimensions are equally important. This statement implies that respondents merely count up the attribute level positions instead of trading-off attributes in the DCE, or focusing mostly on the level positions, irrespective of attribute, in the BWS format.

Third, revealed cognitive burden regarding choice consistency and (simplifying) choice behavior was assessed based on the actual choices of respondents. This included calculating the proportion of respondents providing the same answers to the twice repeated choice task. For the BWS arm, a consistent response was defined as providing the same answer for either best or worst option, following Krucien et al. (2017). Furthermore, we estimated a lexicographic score, which provides information on trading between attribute levels and dominant choice behavior. This score was obtained also following an approach applied by Krucien et al. (2017): First, the proportion of choices based on one attribute on an individual level was calculated. Assuming respondents exhibit dominant preferences for an attribute given proportions above 90% (DCE) and 50% (BWS), the lexicographic score was obtained by calculating the proportion of respondents with such preferences.

To test the impact of color coding on the choice behavior and strategies in the BWS study arms, the shares of responses based on top and bottom levels of the WOOP dimensions were calculated. Additionally, results from the second repeated choice task, where the intensity color coding was reversed, was used to assess the dependence on the color scheme.

Statistical significance was assessed using Wilcoxon-rank sum tests for the Likert scale data (de Winter and Dodou, 2010) and chi-squared tests or Fisher exact tests for proportions. A significance level of 10% was used throughout the analysis. Stata 15 (StataCorp 2017) was used for all calculations.

8.3 Results

8.3.1 Sample characteristics, dropouts, and completion time

A total of 477 participants successfully started with the experiment and were randomly allocated to the three study arms. No respondent dropped out in study arm 1 (DCE). One of the three dropouts in study arm 2 (BWSc) occurred during the choice tasks and two afterwards. Of the five respondents dropping out in study arm 3 (BWS), four dropouts occurred during answering the BWS tasks and one at a later stage. Fisher exact tests indicated that the difference in total drop-out rates was significantly lower in study arm 1 compared to study arm 3 (0% vs. 3.2%, p-value = 0.029). The difference to study arm 2 was not significant (0% vs. 1.9%, p-value = 0.248).

The characteristics of the remaining sample, split by study arm, are shown in Table 1. The randomization led to well-balanced samples regarding most sociodemographic aspects, health status (EQ-5D-5L) and well-being (WOOP). 63.7% of the overall sample was younger than 70 years, 34.6% was aged between 70 and 79 years, and 1.7% were aged 80 years and above with 87 years as the maximum age observed.

The average time it took respondents to complete all 13 choice tasks was 6.0 minutes (SD 3.1) for the DCE tasks, 7.6 minutes for the color coded BWS tasks (SD 4.9) and 7.2 minutes for the standard BWS tasks (SD 4.6). T-tests indicated that choice task completion time was significantly lower for the DCE tasks compared to the two sets of BWS tasks ($p<0.001$ and $p= 0.007$). Fig. 4 plots the mean and median completion times for each choice task separated for each study arm. Differences were most pronounced in the beginning with choice task completion time following a downward trend, likely resulting from learning effects. Finding large differences in mean, but moderate in median answering time in the beginning indicates that some respondents found it particularly difficult to work with and understand the BWS question format compared to the DCE format. On aggregate, respondents in study arm 1 answered each choice task faster compared to the BWS study arms, except for one choice task. Differences within the two BWS study arms were less pronounced with the notable exception of choice task 13, where the intensity color coding was reversed (e.g. light blue corresponded to the worst level and deep purple to the best).

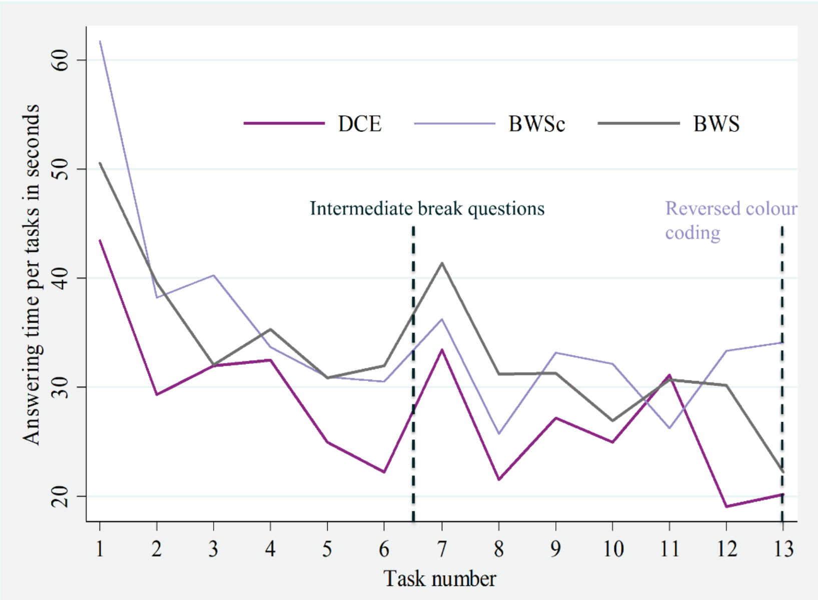
Table 1 Main characteristics of analysis sample per study arm.

	DCE (1)	BWSc (2)	BWS (3)
Age in years	69.3	69.1	68.9
Female (%)	0.65	0.60	0.62
Years of education	16.1	15.8	15.8
Country of residence: UK (ref. U.S.) (%)	0.57	0.54	0.52
Employed (%)	0.33	0.29	0.28
EQ-5D-5L utilities (0-1)	0.83	0.82	0.82
WOOP (Sum score rescaled to 0-1)	0.81	0.79	0.82
Number of completes (N)	159	158	152

Note: EQ-5D-5L tariff from (Devlin et al., 2018).

8.3.2 Self-reported cognitive burden of tasks and number of choice tasks

Mean response scores of the three blocks of debriefing questions and results from significance tests comparing the mean scores across study arms are shown in Table 2. DCE choice tasks appeared to be superior in terms of clarity of the tasks and whether tasks were comprehensible from the beginning. Respondents found the presented states easier to image in the BWS tasks, which admittedly confronted participants only with one well-being state instead of two in the DCE. Color coded BWS choice tasks were evaluated to be less clear than non-color coded BWS tasks.



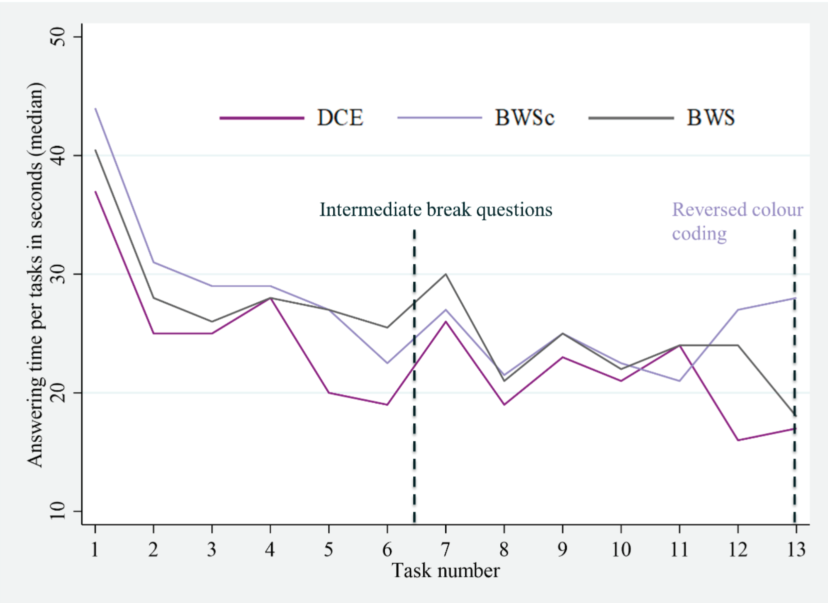


Fig. 4. Mean and median completion times per choice task within each study arm

Results from the second block of questions indicated that participants from the DCE study arm found the number of choice tasks easier to manage, were more able to stay concentrated over all choice tasks, and could have answered more tasks, compared to the BWS study arms, with most differences being statistically significant. Color coding the BWS tasks appeared to have a positive effect on the number of choice tasks participants could handle.

Table 2 Mean response score of cognitive debriefing questions.

Question on Likert scale from 1 to 5 (5=strongly agree)	DCE (1)	BWSc (2)	BWS (3)
Self-reported cognitive burden			
The choice tasks were clear	$4.45^{\dagger ALL}$	$4.11^{\dagger ALL}$	$4.25^{\dagger ALL}$
I could easily choose between the alternatives	3.55	3.65	3.62
I fully understood the choice tasks from the beginning	$4.75^{\dagger ALL}$	$4.26^{\dagger 1}$	$4.36^{\dagger 1}$
The tasks got easier after answering several	3.77	3.87	3.84
I found some of the presented states difficult to imagine	$3.43^{\dagger 3}$	$2.97^{\dagger 1}$	$2.84^{\dagger 1}$
Number of choice tasks			
The number of choice tasks was manageable	$4.64^{\dagger 3}$	4.54	$4.50^{\dagger 1}$
It was difficult to stay concentrated over all choice tasks	$1.72^{\dagger 3}$	1.94	$1.92^{\dagger 1}$
I could have answered more choice tasks	$4.07^{\dagger ALL}$	$3.91^{\dagger ALL}$	$3.66^{\dagger ALL}$
Answering another block of six 6 choice tasks would be manageable	$4.43^{\dagger ALL}$	$4.19^{\dagger 1}$	$4.18^{\dagger 1}$
Choice strategies			
I compared all dimensions/items before making my choice	4.72	4.77	4.79
I decided all dimensions/items are equally important	$2.86^{\dagger 3}$	3.00	$3.20^{\dagger 1}$
I always used the same 1 or 2 well-being dimensions to make my choice	$3.04^{\dagger ALL}$	$2.65^{\dagger 1}$	$2.57^{\dagger 1}$

Note: † $p < 0.10$ of Wilcoxon rank-sum test comparing study arms 1, 2, and 3.

8.3.3 Choice strategies and choice behavior

Most respondents strongly agreed with the statement that they compared all dimensions/items before making their choices, with no significant differences between study arms (Table 2). There were mixed results concerning the use of simplifying choice heuristics or strategies comparing DCE and BWS study arms. While DCE participants agreed to a lesser extent that they decided that all dimensions/items are equally important, they also reported to a larger degree to having based their decisions on the same 1 or 2 well-being dimensions, which implies some level of attribute non-attendance.

Table 3 lists results for the analysis of choice behavior. The lexicographic score (see section 2.5), was significantly lower in DCE respondents, indicating more trading and less dominant choice behavior. In the DCE, dominant preferences were observed only for the physical health attribute. In the BWS, such behavior was also observed for the mental health and making ends meet attributes, with physical health still being the most prevalent one.

In the DCE study arm, 4.4% of respondents did not provide the same answer to the repeated choice task, when it appeared again for the first time (position 2 and 8), with the same color code. When it was repeated again as the last choice task, that share was 2.5%. Up to 20% of respondents did not provide either the same best or worst answer in the repeated BWS tasks.[xvi] When defining consistency as providing the same answer to both best and worst, this share increased to around 60%. There were no significant differences between BWS study arms regarding the choice consistency of the first repeated instance. Almost half of respondents did not provide a consistent best or worst answer to the repeated BWS choice task, where the intensity color coding was reversed (position 13). This share was 72.8% when defining consistency in terms of selecting the same best and worst items.

We further calculated the percentage of best and worst answers based on either the top and bottom levels of the WOOP dimensions on individual level and aggregated that by taking the average. The average share was between 60 and 75%, with higher values observed for the color coded BWS tasks (significant difference for ‘best’).

[xvi] It has to be acknowledged, though that the likelihood of providing the same answer by chance alone is larger for DCE choice tasks (50%).

Table 3 Revealed choice behavior.

	DCE (1)	BWSc (2)	BWS (3)
Non-trading or dominant choice behavior			
Lexicographic score	28.9%$^{\dagger ALL}$	79.1%	80.1%
Choice consistency			
% failed a consistent response to repeated choice task (1st)¶	4.4%$^{\dagger ALL}$	19.6%$^{\dagger 1}$	17.8%$^{\dagger 1}$
% failed a consistent response to repeated choice task (2nd)¶	2.5%$^{\dagger 3}$	46.8%§	19.1%$^{\dagger 1}$
% who did not provide same answer for best and worst (1st)		58.9%	61.2%
% who did not provide same answer for best and worst (2nd)		72.8%§	60.5%
Focus on top and bottom levels			
Mean individual % of choosing level 1 as best		70.5%$^{\dagger 3}$	59.9%$^{\dagger 2}$
Mean individual % of choosing level 5 as worst		76.3%	69.4%

Note: † $p < 0.10$ of chi-squared tests comparing study arms 1, 2, and 3 (if applicable). ¶For BWS defined as providing either the same best or worst answer. §Choice task with intensity colour coding being reversed.

8.4. Discussion

To assess the cognitive burden of different types of choice tasks for valuing well-being states for quality-of-life measures in older people, a randomized experiment was conducted, allocating respondents to either a DCE, a color coded BWS, or a regular BWS format using an online setting. Our study contributes to the literature by providing empirical evidence on 1) whether DCE or BWS choice tasks are associated with lower cognitive burden in the context of health or well-being state valuation in an older population sample, and 2) whether color coding of BWS tasks affects cognitive burden and to a lesser extent validity of BWS experiments.

Finding a lower drop-out rate and lower choice task completion time in the DCE study arm compared to the BWS study arms implies that, for older people, DCE choice tasks are less tiring and faster to complete than BWS tasks. Lower completion time was also observed by van Dijk et al. (2016). In terms of self-reported measures, our results indicate that the DCE tasks also were perceived as less cognitively burdensome, and that a higher number of DCE choice tasks was regarded as more acceptable than was a higher number of BWS tasks. The former has also been reported in related studies in different contexts (Whitty and Oliveira Gonçalves, 2018). The latter is especially relevant to consider when thinking about the number of choices per respondent, and hence the required sample size, when selecting DCE or BWS format. Finding lower cognitive burden associated with DCE tasks compared to BWS tasks, in general, is at odds with what has been reported by Netten et al. (2012). They also compared cognitive burden of DCE and BWS tasks for valuing a large descriptive system of a quality of life instrument, but the design of their study was fairly different. The authors used cognitive interviewing, a qualitative approach, in a small sample (N=30), split the DCE task into two parts to reduce the difficulty of the task and showed both DCE and BWS tasks to respondents.[xvii] Whether the difference in findings relates to the differences in design of the studies, is difficult to say.

In terms of (simplifying) choice strategies and choice behavior, which co-occur with larger cognitive burden, our results are mixed regarding the self-reported behavior, and less clear cut. We did observe a considerably higher choice consistency and lower degrees of dominant choice behavior for DCE respondents, with their measurement to some degree accommodating for the methodological differences. However, these results may relate more to artefacts of the type of choice task and may be unrelated to cognitive burden. As stated also by Whitty et al. (2017), the probability of answering consistently to a DCE task by pure chance is already 50%. With nine dimensions this probability is much lower (22%) for the BWS task (defined as providing either the same best or worst answer). Nevertheless, finding that around 60% of BWS respondents did not provide the same best and worst answers when

[xvii] Although it does not become clear from the paper, whether respondents had to answer full sets of choice tasks or only one task per method.

a choice was repeated for the first and the second time, is somewhat worrisome on its own. A higher degree of trading and lower degrees of dominant choice behavior in DCEs were also reported in the related literature before (Krucien et al., 2017; Whitty et al., 2014) with a similar caveat as for analyzing choice consistency.

Comparing color coded with non-color coded BWS, we found a similar drop-out rate for both tasks (1.9% and 3.2%, respectively). In the study by Jonker et al. (2018) (study arms 1 and 2), color coding of the DCE tasks decreased the dropout rate from 13.9% to 9.8%. Further results from the same study set up showed that color coding alone did not lead to differences with respect to the self-reported cognitive debriefing questions (Jonker et al., 2019). Our results for BWS regarding these questions are mixed. While participants of the color coded BWS on average agreed to a higher extent that they could have answered more choice tasks, the non-color coded BWS choice tasks appeared to have been clearer to respondents. Given no conclusive evidence on cognitive burden, and the fact that the color coding increased the already high focus on top and bottom levels of the quality-of-life instrument in the BWS tasks, color coding BWS cannot be recommended for health or well-being state valuation studies among older people.

The overall implications of our analysis must be interpreted considering several limitations. First, the rather small sample size did not provide us with enough statistical power to be able to use several blocks of choice tasks, which then also would have allowed us to estimate DCE and BWS models. During the design stage, we aimed for 150 respondents per study arm due to the small overall pool of individuals aged 65 on online platforms. While the choice sets were created according to standard design methodology, it could be the case that either of the two choice sets is more difficult to answer in general, irrespective of choice task format, due to smaller utility difference within the shown profiles. As utility weights for the WOOP are not available yet, it was not possible to account for that in the selection of choice set. This risk could have been reduced if multiple blocks would have been used. A second, related, limitation is that DCE and BWS models could not be estimated, which prevented us from analyzing the actual choices people made. Testing for choice consistency or overall noise in the data would have given us an indication on the quality of the responses However, such a comparison between DCE and BWS responses would have come with additional limitations.

In terms of the generalizability of our results, we need to acknowledge the following: Our study was conducted in an online setting, with respondents from an online panel. As certain subpopulations with varying levels of cognitive abilities may self-select into such panels (especially in older ages), the representativity to the general population aged 65 and above may be limited. However, the purpose of our study was to provide an indication of cognitive burden of different methods *specifically* using respondents from online panels, which by now are the most frequently used sampling formats for these types of analyses (Mulhern et al., 2018). Therefore, our results should only be generalized to similar online settings. Our sample likely was on the upper end of the spectrum of cognitive abilities of people aged 65 and above (highly educated and rather healthy, see Table 1). It is not certain, whether our conclusions would be the same in a sample with average or low levels of cognitive abilities, as we did not measure cognitive abilities directly. However, using years of education as an imperfect proxy for overall cognitive abilities, we could not observe an education, and therefore cognitive ability, gradient in our results (i.e. the direction of our results remained stable, when splitting our sample into a lower and a higher educated group). To increase the representativeness of the sample in a full-scale valuation study among the elderly using online panels, it will be necessary to implement further age stratification by setting appropriate age group quotas.

As for the generalizability towards other online panels, the following limitation applies: Per online platform rule, the recruitment of respondents involved a monetary compensation which is rather high compared to standard online panels, and which can be reduced if the researcher is not satisfied with the quality of responses. While this is a good thing for respondents and their motivation, this led to very low dropout rates and could have also affected other parts of the analysis Another caveat of our analysis is that the applicability of our results to the comparison of DCEs without overlap and colour coding, and BWS is limited. However, the use of level overlap in similar DCEs as strategy to reduce task complexity seems to be increasing (e.g. King et al., 2018; Mulhern et al., 2019). Not really a limitation, but important to note in terms of cognitive burden is the following: In the DCE setup, it was possible and logical to reduce the level descriptions compared to the full level text in the BWS, as the attributes were already included on the left side of the task (Fig. 3). This may also have contributed to DCE tasks being perceived to be easier to handle.

8.5 Conclusions

Overall, we found evidence that level overlapped, and color coded DCE choice tasks are less cognitively burdensome than BWS choice tasks, in a complex health (or, here, well-being) state valuation exercise among older people in an online setting. This has implications for future valuation studies, especially since the complexity of the measures to be valued seems to increase when moving from health-related to overall quality of life; see, for instance, the WOOP (Appendix A), the current plans of the E-QALY project (https://scharr.dept.shef.ac.uk/e-qaly/), or another ongoing study developing a quality of life measure for older people (Ratcliffe et al., 2019). Cognitive burden should be an important factor in deciding about which method to choose for valuing such descriptive systems, but at the same time, statistical and theoretical aspects need to be considered as well. Although our results may not be easily generalizable to other topics of study within or outside health care and to other study populations, our analysis may at least serve as a good example of how to assess cognitive burden associated with different types of choice experiments.

References

Arendts, G., Jan, S., Beck, M.J., Howard, K., 2017. Preferences for the emergency department or alternatives for older people in aged care: A discrete choice experiment. Age Ageing 46, 124–129. https://doi.org/10.1093/ageing/afw163

Cheung, K.L., Wijnen, B.F.M., Hollin, I.L., Janssen, E.M., Bridges, J.F., Evers, S.M.A.A., Hiligsmann, M., 2016. Using Best–Worst Scaling to Investigate Preferences in Health Care. Pharmacoeconomics 34, 1195–1209. https://doi.org/10.1007/s40273-016-0429-5

Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J., Peters, T.J., 2008. Valuing the ICECAP capability index for older people. Soc. Sci. Med. 67, 874–882. https://doi.org/10.1016/j.socscimed.2008.05.015

de Meijer, C., Wouterse, B., Polder, J., Koopmanschap, M., 2013. The effect of population aging on health expenditure growth: A critical review. Eur. J. Ageing. 10, 353–361. https://doi.org/10.1007/s10433-013-0280-x

de Winter, J., Dodou, D., 2010. Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). Pract. Assessment, Res. Eval. 15. https://doi.org/https://doi.org/10.7275/bj1p-ts64

Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B., van Hout, B., 2018. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 27, 7–22. https://doi.org/10.1002/hec.3564

Flynn, T.N., 2010. Valuing citizen and patient preferences in health: Recent developments in three types of best-worst scaling. Expert Rev. Pharmacoeconomics Outcomes Res. 10, 259–267. https://doi.org/10.1586/erp.10.29

Flynn, T.N., Huynh, E., Peters, T.J., Al-Janabi, H., Clemens, S., Moody, A., Coast, J., 2015. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. Heal. Econ. (United Kingdom) 24, 258–269. https://doi.org/10.1002/hec.3014

Flynn, T.N., Marley, A.A.J., 2014. Best-worst scaling: theory and methods, in: Handbook of Choice Modelling. Edward Elgar Publishing, pp. 178–201. https://doi.org/10.4337/9781781003152.00014

Franco, M.R., Howard, K., Sherrington, C., Rose, J., Ferreira, P.H., Ferreira, M.L., 2016. Smallest worthwhile effect of exercise programs to prevent falls among older people: estimates from benefit-harm trade-off and discrete choice methods. Age Ageing 45, 806–812. https://doi.org/10.1093/ageing/afw110

Hackert, M.Q.N., Brouwer, W.B.F., Hoefman, R.J., van Exel, J., 2019. Views of older people in the Netherlands on wellbeing: A Q-methodology study. Soc. Sci. Med. 240, 112535. https://doi.org/10.1016/j.socscimed.2019.112535

Jonker, M.F., Attema, A.E., Donkers, B., Stolk, E.A., Versteegh, M.M., 2017. Are Health State Valuations from the General Public Biased? A Test of Health State Reference Dependency Using Self-assessed Health and an Efficient Discrete Choice Experiment. Heal. Econ. (United Kingdom) 26, 1534–1547. https://doi.org/10.1002/hec.3445

Jonker, M.F., Donkers, B., de Bekker-Grob, E., Stolk, E.A., 2019. Attribute level overlap (and colour coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Heal. Econ. (United Kingdom) 28, 350–363. https://doi.org/10.1002/hec.3846

Jonker, M.F., Donkers, B., de Bekker-Grob, E.W., Stolk, E.A., 2018. Effect of Level Overlap and Colour Coding on Attribute Non-Attendance in Discrete Choice Experiments. Value Heal. 21, 767–771. https://doi.org/10.1016/j.jval.2017.10.002

King, M.T., Viney, Rosalie, Simon Pickard, A., Rowen, Donna, Aaronson, N.K., Brazier, J.E., Cella, David, Costa, D.S.J., Fayers, P.M., Kemmler, Georg, McTaggart-Cowen, H., Mercieca-Bebber, Rebecca, Peacock, S., Street, D.J., Young, T.A., Norman, Richard, Aaronson, N., Brazier, J., Cella, D., Costa, D., Fayers, P., Grimison, P., Janda, M., Kemmler, G., King (Chair), M., McTaggart-Cowan, H., Mercieca-Bebber, R., Norman, R., Pickard, S., Rowen, D., Velikova, G., Viney, R., Street, D., Young, T., 2018. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. Pharmacoeconomics 36, 225–238. https://doi.org/10.1007/s40273-017-0582-5

Krucien, N., Watson, V., Ryan, M., 2017. Is Best–Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments. Heal. Econ. (United Kingdom) 26, e1–e16. https://doi.org/10.1002/hec.3459

Louviere J.L., 2004. Random utility theory-based stated preference elicitation methods: applications in health economics with special reference to combining sources of preference data. Working Paper at Center for the Study of Choice No. 04-001

Maddala, T., Phillips, K.A., Johnson, F.R., 2003. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ. 12, 1035–1047. https://doi.org/10.1002/hec.798

Makai, P., Brouwer, W.B.F., Koopmanschap, M.A., Stolk, E.A., Nieboer, A.P., 2014. Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. Soc. Sci. Med.

102, 83–93. https://doi.org/10.1016/j.socscimed.2013.11.050
Milte, R., Ratcliffe, J., Chen, G., Lancsar, E., Miller, M., Crotty, M., 2014. Cognitive overload? An exploration of the potential impact of cognitive functioning in discrete choice experiments with older people in health care. Value Heal. 17, 655–659. https://doi.org/10.1016/j.jval.2014.05.005
Mulhern, B., Norman, R., De Abreu Lourenco, R., Malley, J., Street, D., Viney, R., 2019. Investigating the relative value of health and social care related quality of life using a discrete choice experiment. Soc. Sci. Med. 233, 28–37. https://doi.org/10.1016/j.socscimed.2019.05.032
Mulhern, B., Norman, R., Street, D.J., Viney, R., 2018. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. Pharmacoeconomics. https://doi.org/10.1007/s40273-018-0714-6
Neumann, P.J., Sanders, G.D., Russell, L.B., Siegel, J.E., Ganiats, T.G., 2016. Cost effectiveness in health and medicine. Oxford University Press, New York.
Palan, S., Schitter, C., 2018. Prolific.ac—A subject pool for online experiments. J. Behav. Exp. Financ. 17, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004
Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., Brazier, J.E., 2011. Best-worst scaling vs. discrete choice experiments: An empirical comparison using social care data. Soc. Sci. Med. 72, 1717–1727. https://doi.org/10.1016/j.socscimed.2011.03.027
Ratcliffe, J., Cameron, I., Lancsar, E., Walker, R., Milte, R., Hutchinson, C.L., Swaffer, K., Parker, S., 2019. Developing a new quality of life instrument with older people for economic evaluation in aged care: study protocol. BMJ Open 9, e028647. https://doi.org/10.1136/bmjopen-2018-028647
Ryan, M., Gerard, K., Amaya-Amaya, M. (Eds.), 2008. Using Discrete Choice Experiments to Value Health and Health Care, The Economics of Non-Market Goods and Resources. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-1-4020-5753-3
Soekhai, V., de Bekker-Grob, E.W., Ellis, A., Vass, C.M., 2018. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. https://doi.org/10.1007/s40273-018-0734-2
van Dijk, J.D., Groothuis-Oudshoorn, C.G.M., Marshall, D.A., IJzerman, M.J., 2016. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. Value Heal. 19, 316–322. https://doi.org/10.1016/j.jval.2015.12.020
Whitty, J.A., Oliveira Gonçalves, A.S., 2018. A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best–Worst Scaling for Eliciting Preferences in Healthcare. Patient 11, 301–317. https://doi.org/10.1007/s40271-017-0288-y
Whitty, J.A., Walker, R., Golenko, X., Ratcliffe, J., 2014. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. PLoS One 9. https://doi.org/10.1371/journal.pone.0090635
Yao, R.T., Scarpa, R., Rose, J.M., Turner, J.A., 2015. Experimental Design Criteria and Their Behavioural Efficiency: An Evaluation in the Field. Environ. Resour. Econ. 62, 433–455. https://doi.org/10.1007/s10640-014-9823-7

Appendices

Appendix A

Well-being of Older People (WOOP) instrument

For each section, select the description that is most appropriate for you today.

Physical health

Consider physical conditions or ailments and other physical impairments that affect your daily functioning.

☐ I have no problems with my physical health
☐ I have slight problems with my physical health
☐ I have moderate problems with my physical health
☐ I have severe problems with my physical health
☐ I have very severe problems with my physical health

Mental health

Consider problems with your ability to think, anxiety, depression and other mental impairments that affect your daily functioning.

☐ I have no problems with my mental health
☐ I have slight problems with my mental health
☐ I have moderate problems with my mental health
☐ I have severe problems with my mental health
☐ I have very severe problems with my mental health

Social life

Consider your relationship with your partner, family or other people who are important to you. This concerns the amount and quality of the contact you have.

☐ I'm very satisfied with my social life
☐ I'm satisfied with my social life
☐ I'm reasonably satisfied with my social life
☐ I'm dissatisfied with my social life
☐ I'm very dissatisfied with my social life

Receive support

Everyone needs help or support sometimes. Consider practical or emotional support, for example from your partner, family, friends, neighbours, volunteers or professionals. This concerns being able to count on support when you need it, as well as the quality of the support.

☐ I'm very satisfied with the support I get, when needed
☐ I'm satisfied with the support I get, when needed
☐ I'm reasonably satisfied with the support I get, when needed
☐ I'm dissatisfied with the support I get, when needed
☐ I'm very dissatisfied with the support I get, when needed

Acceptance and resilience

Consider your acceptance of your current circumstances and your ability to adapt to changes to these, whether or not with support of your religion or belief.

- ☐ I’m very able to deal with my circumstances and changes to these
- ☐ I’m able to deal with my circumstances and changes to these
- ☐ I’m reasonably able to deal with my circumstances and changes to these
- ☐ I’m not able to deal with my circumstances and changes to these
- ☐ I’m not at all able to deal with my circumstances and changes to these

Feeling useful

Consider meaning something to others, your environment or a good cause.

- ☐ I feel very useful
- ☐ I feel useful
- ☐ I feel reasonably useful
- ☐ I do not feel useful
- ☐ I do not feel at all useful

Independence

Consider being able to make your own choices or doing the activities that you find important.

- ☐ I feel very independent
- ☐ I feel independent
- ☐ I feel reasonably independent
- ☐ I feel dependent
- ☐ I feel very dependent

Making ends meet

Consider having enough money to meet your daily needs and having no money worries.

- ☐ I’m more than able to make ends meet
- ☐ I’m able to make ends meet
- ☐ I’m reasonably able to make ends meet
- ☐ I’m not able to make ends meet
- ☐ I’m not at all able to make ends meet

Living situation

Consider living in a house or neighbourhood you like.

- ☐ I’m very satisfied with my living arrangements
- ☐ I’m satisfied with my living arrangements
- ☐ I’m reasonably satisfied with my living arrangements
- ☐ I’m dissatisfied with my living arrangements
- ☐ I’m very dissatisfied with my living arrangements

Appendix B

Repeated choice tasks

Which of the described well-being states do you prefer, A or B? *(2 of 6)*

	A	B
Physical health	**Moderate** problems	**Moderate** problems
Mental health	**Moderate** problems	**Severe** problems
Social contacts	**Very dissatisfied**	**Reasonably satisfied**
Receiving support	**Very satisfied**	**Very dissatisfied**
Acceptance	**Almost unable** to cope	**Almost unable** to cope
Feeling useful	Feeling **very** useful	Feeling **very** useful
Independency	Feeling **very** independent	Feeling **very** independent
Making ends meet	**Well** able to meet ends	**Well** able to meet ends
Living situation	**Satisfied**	**Very dissatisfied**
	◯	◯

- Positive aspects are **light blue** and negative aspects are **darker purple**
- Put the cursor above the underlined items for descriptions

Imagine living in this well-being state and select which aspect you would most prefer, and which aspect you would least prefer. *(2 of 6)*

Most	Well-being state	Least
◯	I am satisfied with my social contacts	◯
◯	I am dissatisfied with the support I receive	◯
◯	I am almost unable to deal with my circumstances and changes therein	◯
◯	I feel useful	◯
◯	I have very severe problems with my physical health	◯
◯	I have no problems with my mental health	◯
◯	I feel independent	◯
◯	I am very well able to make ends meet	◯
◯	I am satisfied with my living situation	◯

- Positive aspects are **light blue** and negative aspects are **darker purple**
- Put the cursor above the underlined items for descriptions

Appendix C

Utility function for DCE design

The following utility function was optimized in Ngene, where i indicates the respondent and j the well-being profile:

$$\begin{aligned} U_{ij} = PH_{ij}\beta_{PH} + MH_{ij}\beta_{MH} + SOC_{ij}\beta_{SOC} + SUP_{ij}\beta_{SUP} + ACC_{ij}\beta_{ACC} \\ + USE_{ij}\beta_{USE} + IND_{ij}\beta_{IND} + MEM_{ij}\beta_{MEM} + LIV_{ij}\beta_{LIV} + \varepsilon_{ij} \end{aligned} \quad (1)$$

PH, MH, SOC, SUP, ACC, USE, IND, MEM, and LIV symbolize vectors of the levels of the WOOP instrument (Appendix A). The betas represent vectors of four parameters each, which model the utility associated with each of the levels of the nine dimensions of the WOOP compared to the lowest level in each dimension.

Chapter 9

Comparing DCE and BWS-2 Outcomes: an application to Neuromuscular Disease treatment

Soekhai V, Donkers B, Viberg Johansson J, Jimenez-Moreno C, Pinto CA, de Wit GA, de Bekker-Grob EW

Submitted

Abstract

Objectives: An increasingly popular method to elicit patient preferences is case 2 best-worst scaling (BWS-2), next to the popular discrete choice experiments (DCEs). Since BWS-2 potentially has a lower cognitive burden compared to DCE, comparing these methods within a patient sample where cognitive burden is relevant may lead to new insights. The aim of this study was therefore to compare treatment preference weights and relative importance scores (RIS) obtained with each method within neuromuscular diseases (NMD) patients.

Methods: NMD patients completed an online survey at two different moments in time, completing one method per occasion. Patients were randomly assigned to either first DCE or BWS-2. Attributes included: muscle strength, energy endurance, balance, cognition, chance of blurry vision, and chance of liver damage. Multinomial logit (MNL) was used to calculate overall RIS and latent class logit (LC) was used to estimate heterogeneous preference weights and to calculate the RIS of the attributes for each latent class.

Results: A total of 140 patients completed DCE and BWS-2 and were included for analyses. Overall RIS showed differences in attribute importance rankings between DCE and BWS-2. Latent class analyses indicated three latent classes for both methods, with a specific class in both DCE and BWS-2 in which (avoiding) liver damage was the most important attribute. Ex-post analyses showed that classes differed in sex, age, level of education and disease status, with patients in the DCE class where liver damage was most important were also more often in the BWS-2 class where this attribute was considered most important. DCE was not found to be more difficult to understand than BWS-2.

Conclusions: This study showed that using different preference elicitation methods leads to different outcomes, both in preference weights as well as in RIS, although latent class analysis revealed similar latent classes between methods. Our results suggest that BWS-2 is the preferred method of choice when dealing with small samples, while DCE may be preferred

when minimizing cognitive burden is key and choice tasks include both benefits and risks. Therefore, careful consideration about method selection is required, while keeping the specific decision context in mind.

9.1 Introduction

There is an emerging consensus that patient preferences should be incorporated within decisions in the medical product lifecycle (MPLC).[1–4] These preferences have become more important for the companies that develop new medical products and for the authorities that assess, regulate, and decide which products are effective, safe, well-tolerated, and cost-effective.[5] Yet, there are still outstanding questions related to which preference methods are best suited for each decision context and there are a lot of different methods that can be used to gain insights into preferences. Studies by for example the Medical Device Innovation Consortium (MDIC)[6] and Soekhai et al.[7] provide an overview of several stated preference methods to elicit these preferences within the MPLC context.

One of the stated preference methods that has become increasingly popular to elicit patient preferences is best-worst scaling (BWS).[8,9] BWS was introduced to obtain more preference information than a discrete choice experiment (DCE) by asking individuals not only to select their best but also their worst option, without a large increase in the cognitive burden of the elicitation task.[8] The literature distinguishes between three types of BWS: object case (case 1 BWS) where attributes (characteristics), profile case (case 2 BWS) where attribute levels (values of characteristics), and multi-profile case (case 3 BWS) where profiles are selected as best and worst.[10] For more details regarding BWS see Louviere et al.[10] Case 2 BWS (hereafter: BWS-2) received much attention in preference literature, since this method is able to uncover attribute level importance, might reduce cognitive burden of the elicitation task by focusing on one profile at a time and is relatively easy to design.[11,12]

Although BWS-2 is being used more frequently in health preference research, it can not yet match the years of expericience and the resulting body of work of DCEs in health preference research.[13,14] In DCEs respondents are presented with multiple-choice tasks including two or more hypothetical alternatives. These alternatives consist of a fixed set of attributes with varying attribute levels between the alternatives and choice tasks. Respondents are then asked to select their preferred alternative in each choice task. For more information about DCEs, see Hensher et al.[15] and Train.[16]

There are few studies investigating differences between DCE and BWS-2 preference study outcomes. Studies from van Dijk (hip replacement surgery)[11], Potoglou et al. (social care preferences)[17] and Severin et al. (priority setting for genetic testing)[18] are examples in which DCE and BWS-2 preferences have been compared. The aim of this study is to compare preference weights and relative importance scores obtained from both methods. In this study we focused on treatment preferences for patients with neuromuscular diseases (NMD), which are rare diseases and often affect the central nervous system (CNS) leading to impaired or reduced cognitive functioning.[19–22] General cognitive deficits have been described in over 60 to 70% of patients and the prevalence and severity depends on the age at onset of the disease. With earlier onset of disease, the cognitive limitations are generally more severe than observed for adult phenotypes, which are classified as those with symptoms first diagnosed ≥20 years of age.[23] Comparing DCE and BWS outcomes in this study context is of interest, as DCEs generally require larger sample sizes, which is challenging for rare disease applications, and NMD patients may have reduced cognitive functioning as the perception is that BWS-2 presents a lower cognitive burden for patients.[24] The latter is related to the fact that previous research showed that BWS-2 requires to frame attributes either all positive or negative (i.e., mixing benefits and risks leads to identifcation problems)[25], while in DCEs combining positive and negative attributes within one choice task is possible, making it cognitively more demanding.

9.2. Methods

9.2.1 Study population

A sample of adult patients with NMD was selected between May and December 2020. Respondents were mostly recruited through patient organizations and patient registries in the UK, USA, Canada, Australia and New Zealand via email, advertisements and newsletters. Informed consent was obtained before the start of the survey. Respondents were included if they were 18 years of age or older, were self-reported as diagnosed with NMD with late onset (established diagnosis or first reported symptoms on or after 20 years of age) and had an active email account to register. Respondents were excluded if they were unable to provide informed consent, complete the online survey, or with reported history of encephalopathy or dementia (as these may have an impact on cognitive skills and ability to complete the survey). This study was approved by the Newcastle University Ethics Committee (Ref: 8840/2018).

9.2.2 Attributes and attributes levels

Potentially relevant attributes and attribute levels for a hypothetical medicinal treatment for NMD patients were selected using a qualitative study for both DCE and BWS-2. The qualitative study included 52 participants who completed in-person semi-structured interviews or participated in focus group discussions. Details regarding these qualitative findings were reported somewhere else.[26,27] Based on this work, six attributes were included in the DCE and BWS-2: muscle strength, energy endurance, balance, cognition, chance of (temporary) blurry vision and chance of (permanent) liver damage. Table 1 presents the attributes and attribute levels for DCE and BWS-2.

Table 1 – Attributes and levels for eliciting preferences with DCE and BWS-2 (including priors for DCE design)

Attributes	Levels		
Muscle strength	**Stays the same** [Ref]	**Improved by half** [0.05,0.15][i] [0.89,0.45][ii]	**Cured** [0.15,0.25][i] [0.95,0.49][ii]
Energy endurance	**Stays the same** [Ref]	**Improved by half** [0.05,0.15][i] [0.60,0.30][ii]	**Cured** [0.15,0.25][i] [0.70,0.36][ii]
Balance	**Stays the same** [Ref]	**Improved by half** [0.05,0.15][i] [0.42,0.21][ii]	**Cured** [0.15,0.25][i] [1.05,0.54][ii]
Cognition	**Stays the same** [Ref]	**Improved by half** [0.05,0.15][i] [0.05,0.61][iii]	**Cured** [0.15,0.25][i] [0.15,0.71][iii]
Chance of (temporary) blurry vision	**1%** [Ref]	**15%** [-0.15,-0.05][i] [-0.59,-0.05][iii]	**30%** [-0.25,-0.15][i] [-0.85,-0.15][iii]
Chance of (permanent) liver damage	**1%** [Ref]	**15%** [-0.15,-0.05][i] [-0.65,0.33][ii]	**30%** [-0.25,-0.15][i] [-1.86,0.95][ii]

i uniformly distributed pilot prior: min-max
ii normally distributed post-pilot updated prior: mean, standard deviation
iii uniformly distributed post-pilot updated prior: min-max

9.2.3 Design of DCE choice tasks

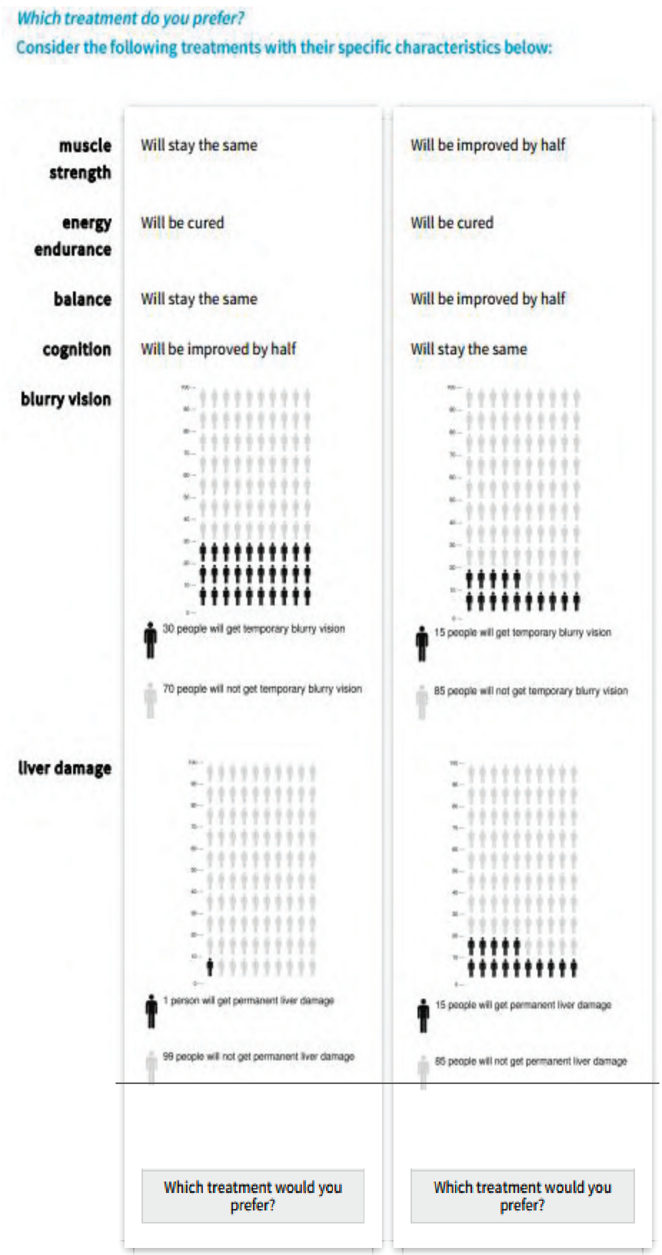
A Bayesian D-efficient design was generated for the DCE, in which the D-efficiency was maximized using Ngene software (Ngene, version 1.2.1).[28] Pilot data from the first 51 respondents were used to update priors and their specific distribution (see Table 1) as well as for further optimization of the design.[28,29] The final DCE design used for the survey included 24 unique choice tasks, which were blocked into two blocks with 12 choice tasks each to reduce cognitive burden for respondents. The alternatives in each choice task were unlabeled and the attribute order was kept constant across all tasks.[30]

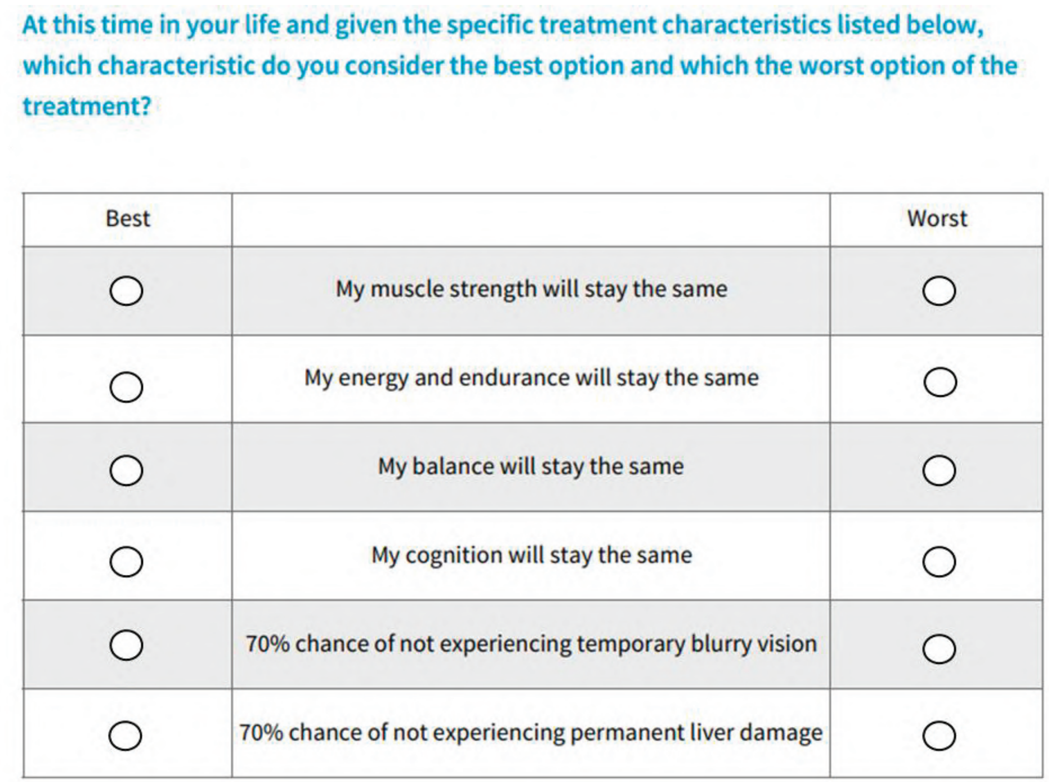
9.2.4 Design of BWS choice tasks

For designing the BWS-2 choice tasks an orthogonal main effect plan (OMEP) experimental design was used. An OMEP enables the independent estimation of preference weights for each attribute level.[10] Based on the number of attributes and levels, the OMEP indicated 18 choice tasks to be included in the experiment.[31] Since the combination of negative and positive attributes in BWS-2 choice tasks can lead to identification problems, negative attributes (i.e. chance of blurry vision and chance of liver damage) were framed positively.[25] This means that for these attributes, attribute levels in Table 1 for BWS-2 included 70%, 85% and 99% chance of not experiencing blurry vision or liver damage. Attribute order was kept constant across all tasks.

9.2.5 Survey design

The survey consisted of several sections. At T=1 this included (1) background questions, such as demographics (age, sex, school or work situation, country of origin), recruitment platform, clinical characteristics (diagnosis and age of diagnosis), disease status and a list of 18 activities along with questions about whether or not these were possible for the patient; (2) a short video introducing the preference task, (3) either BWS-2 (18 choice tasks) or DCE (12 choice tasks) (randomly allocated) and (4) evaluation questions about the ease of understanding and answering, and the usefulness of the video instructions. At T=2, a short video introduced the other preference method and follow-up questions were also included.[26] To minimize cognitive burden, the first set of choice tasks (either DCE or BWS-2) and the second set of choice tasks were administered at different time points, with a two-week period in between. In BWS-2 respondents had to select their best and worst attribute level, while in the DCE respondents were asked about their preferences by choosing between two alternatives. The survey was designed using Lighthouse Studio (Sawtooth Software, version 9.8.1X). Example DCE and BWS-2 choice tasks are shown in Figure 1.





At this time in your life and given the specific treatment characteristics listed below, which characteristic do you consider the best option and which the worst option of the treatment?

Best		Worst
○	My muscle strength will stay the same	○
○	My energy and endurance will stay the same	○
○	My balance will stay the same	○
○	My cognition will stay the same	○
○	70% chance of not experiencing temporary blurry vision	○
○	70% chance of not experiencing permanent liver damage	○

Figure 1 – Example DCE and BWS-2 choice tasks

9.2.6 Statistical analysis

Statistical analyses were performed using data from respondents who completed both BWS-2 and DCE tasks (including respondents from pilot). Following guidance from the literature, as well as our interest in investigating preference heterogeneity, identifying different respondent groups and model fit, a latent class model (LC) was estimated to analyze choice data for both DCE and BWS-2.[10,15] While the standard multinomial logit model (MNL), used as a starting point within this study, assumes that all respondents have identical preferences, LC deals with preference heterogeneity by assuming - based on the choices respondents made - that there are a fixed number of different groups of respondents (i.e., latent classes).[16] Within each group in traditional LC each individual has identical preferences.

With LC, the utility (U) of an alternative for each latent class in both DCE and BWS-2 can be modeled as a linear function of the specific attributes and levels, with

$$U = \sum_{k=1}^{A} \sum_{j=1}^{J_k} \beta_{k,j} X_{k,j} + \varepsilon \qquad \text{eq. 1}$$

where there are A attributes with attribute *k* having J_k attribute levels, with $X_{k,j}$ equal to one if the attribute level j of an attribute k is available in the presented profile, $\beta_{k,j}$ are the utility parameters for the j^{th} levels of attribute k and ε being the random error term representing the unexplained part of utility. LC was programmed using R version 4.0.0 (Apollo package, version 0.0.1) to estimate the utilities for both the DCE and BWS-2 data, as well as for the ex-post descriptive analyses to characterize the latent classes.[32,33] For DCE and BWS-2, “muscle strength stays the same” was selected as reference level (fixed at zero). DCE also required a reference level within each specific attribute. To create a clear interpretation of attribute levels (for attributes muscle strength, energy endurance, balance and cognition), the least attractive attribute levels were used as reference level. For the other attributes the most attractive attribute levels were selected as reference level This means that for muscle strength, energy endurance, balance and cognition preference weights increase when the attribute level value increases, while for the chance of blurry vision and chance of liver damage the preference weights decrease with increasing attribute levels. To facilitate the comparison between DCE and BWS-2, the utility levels relative to the corresponding attribute reference level were also estimated for BWS-2. Relative importance scores (RIS) of attributes were calculated by looking at the maximum utility differences between two attribute levels within each specific attribute and compared between DCE and BWS-2, while outcomes from the evaluation questions for both methods were also analyzed.

9.3. Results

A total of 140 patients completed both the DCE and BWS-2 part of the survey. Responding patients were mostly female (65%) and the median age was 54 (with a range of 23-76). The majority of patients completed a higher (45%) or vocational (34%) education. Most patients reported that they were able to walk without an assistive device (36%), followed by 26% of the patients reporting to be able to walk but relying on an assistive device. A relatively large group of patients (23%) also reported to be able to walk and run without an assistive device (Table 2).

Table 2 – Sample characteristics

Characteristic	DCE & BWS-2
Respondents	140
Sex	
Female	91 (65%)
Age (years)	
Median (range)	54 (23-76)
Highest level of education	
No formal schooling	1 (1%)
Elementary	5 (4%)
Secondary	21 (15%)
Vocational	48 (34%)
Higher	63 (45%)
No answer	2 (1%)
Disease status	
Walk and run without assistive device	32 (23%)
Walk without assistive device	50 (36%)
Walk but rely on assistive device	36 (26%)
Walk but using wheelchair part-time	19 (14%)
Fully rely on wheelchair	3 (1%)

Figure 2 shows the overall (based on MNL) RIS calculations for both DCE and BWS-2. For DCE, (avoiding) liver damage had the highest relative importance, followed by muscle strength, energy endurance, balance, cognition and (avoiding) blurry vision. For BWS-2, a different pattern was observed. Muscle strength had the highest RIS value, followed by energy endurance, balance, liver damage, cognition and blurry vision. Preferences for improving the typical impairments of NMD were similar across methods, with generally a high preference to improve muscle strength, energy and (to a somewhat lesser extent) balance. Zooming in by accounting for preference heterogeneity with LC, Figure 3a and 3b illustrate the relative importance of each attribute for each latent class. Given the sample size, statistical measures of fit and aiming for a meaningful interpretation of the latent classes, a three-class model was superior for both DCE and BWS-2. The DCE latent classes in Figure 3a reveal a group of patients in which avoiding liver damage is by far the most important attribute, while there are also patient groups where improvement of balance and energy endurance are most important. For BWS-2, there is a patient group in which muscle strength is most important, while there is – similar to DCE – a patient group in which liver damage is considered the most important attribute (Figure 3b).

Table 3 presents the estimated LC preference weights for both preference methods. Focusing on the magnitude of these weights, for DCE overall the more attractive levels were preferred above the less attractive levels with most attribute levels being statistically significant. This is however not the case in DCE class 2, in which most attribute levels are not statistically significant and where the utility of 15% chance of liver damage was larger than the utility of 1% chance of liver damage. The largest patient class (47%) was the class of patients in which liver damage was the most important attribute (class 3). For BWS-2, Table 3 shows that most attribute levels were statistically significant. Additionally, all more attractive attribute levels were preferred above the less attractive attribute levels. The largest classes of patients were the classes in which energy endurance (42%) and liver damage (41%) were the most important attributes.

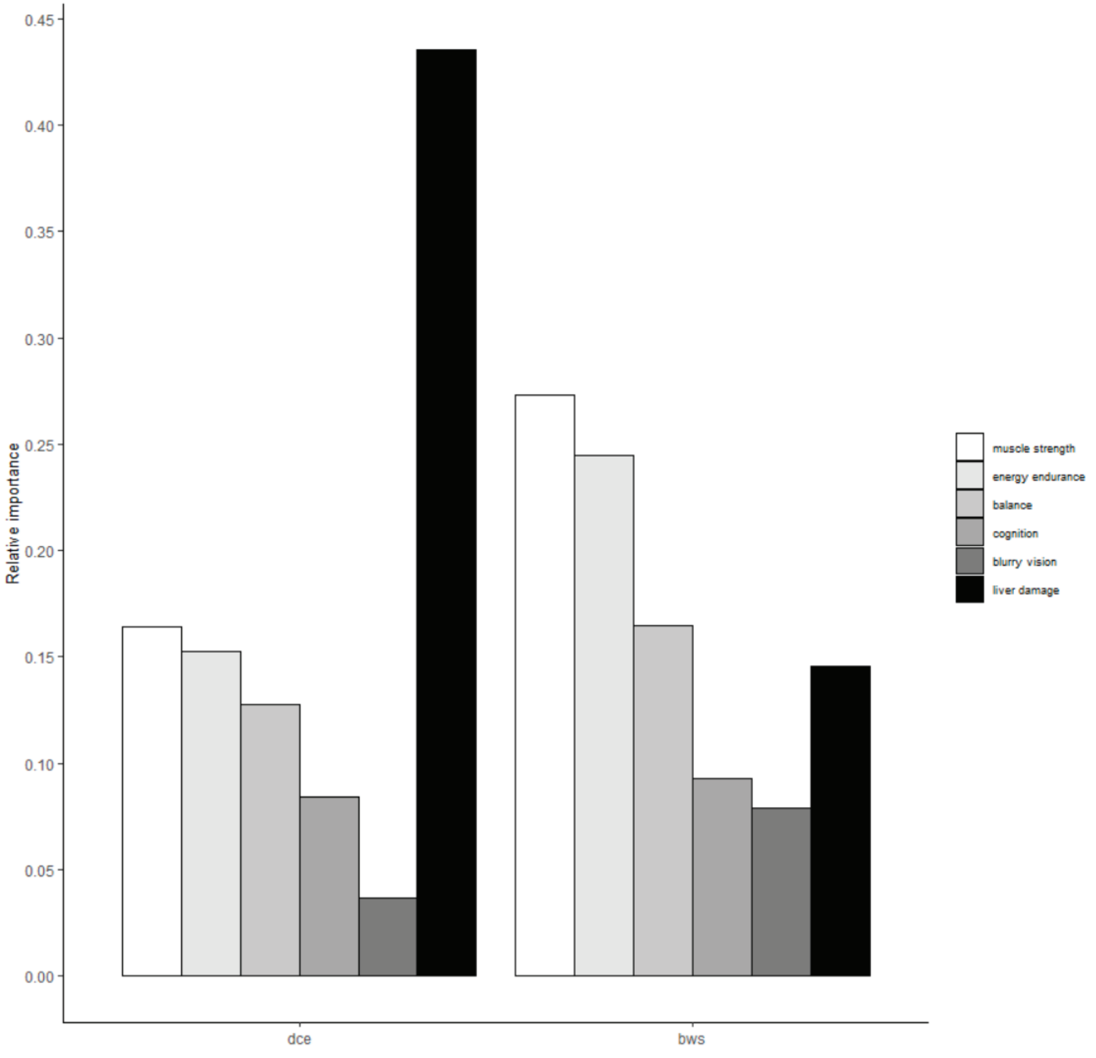
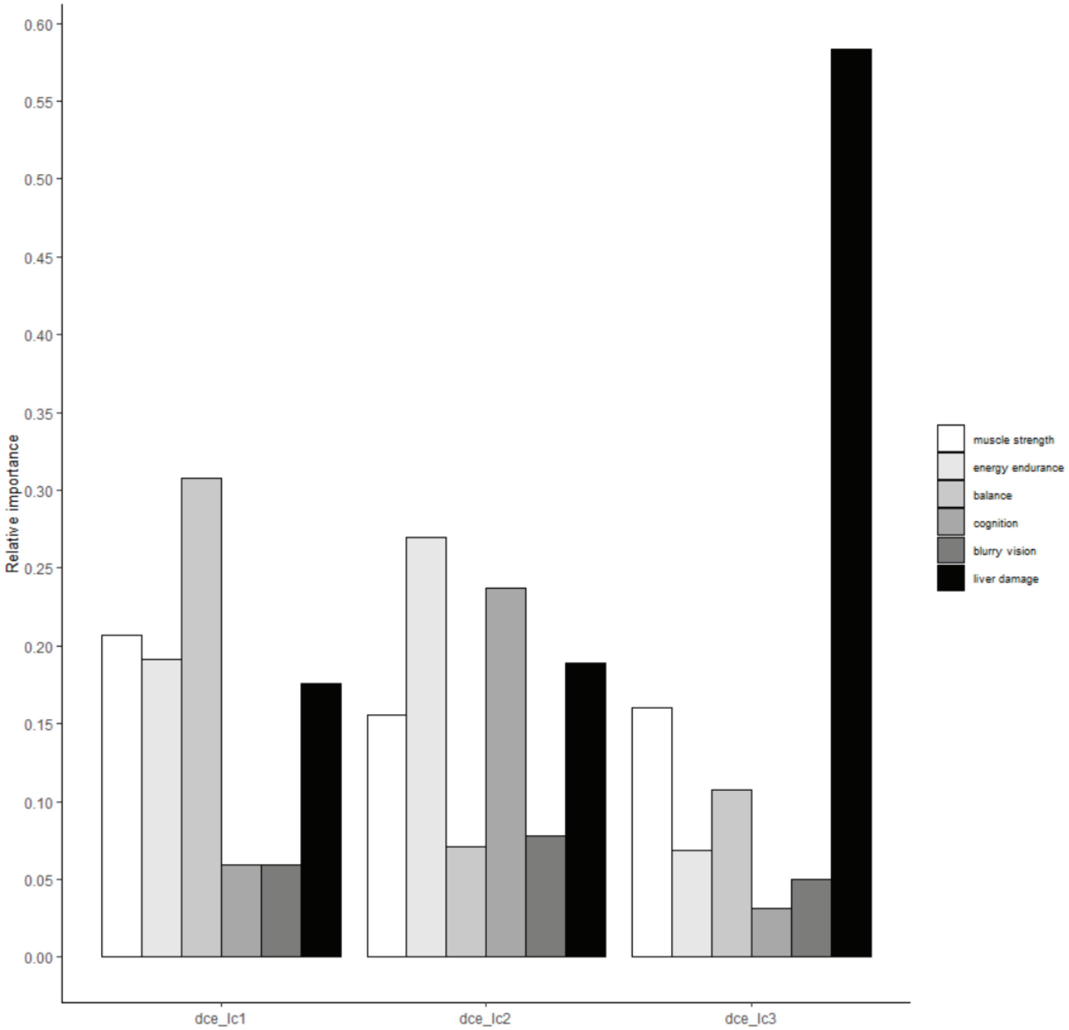
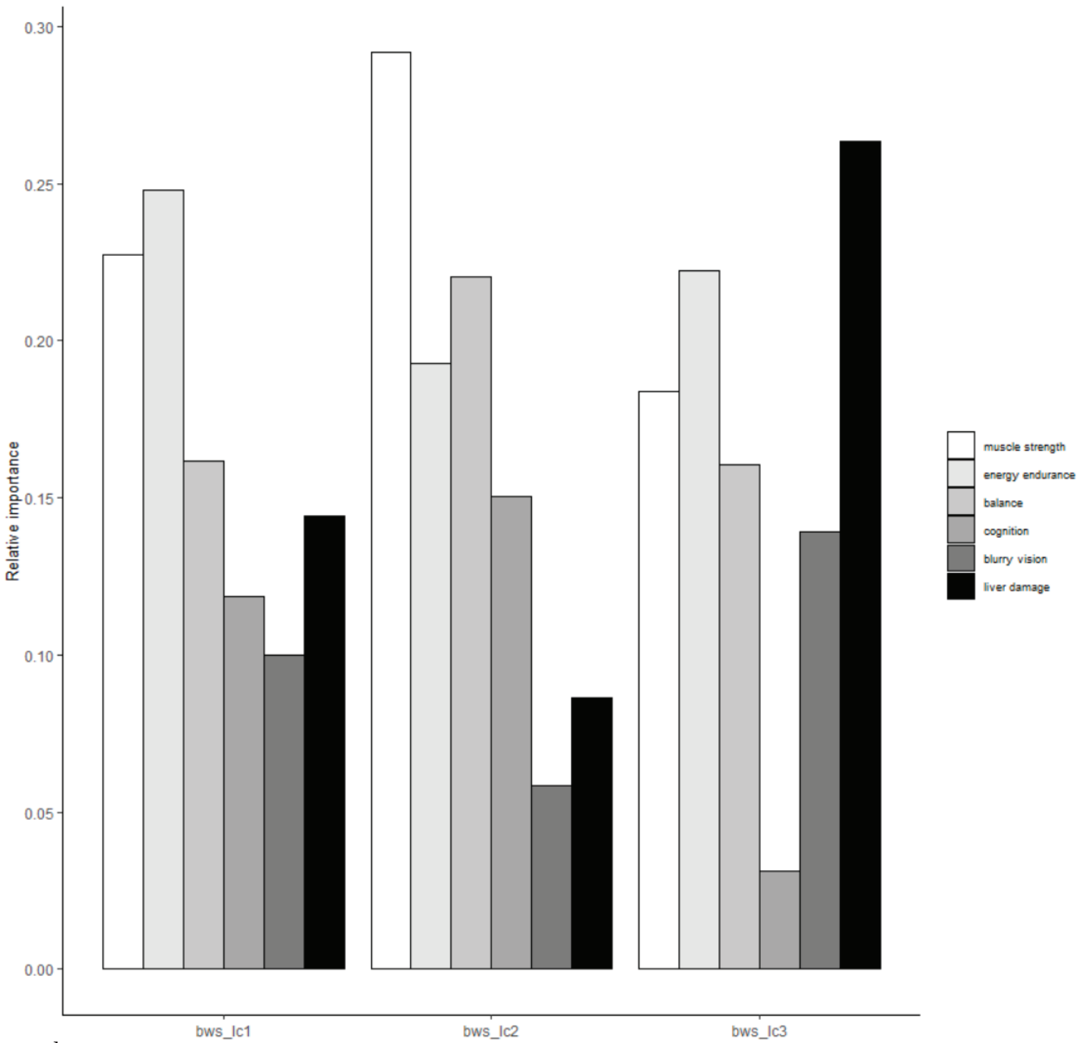


Figure 2 – Overall relative importance score of attributes for DCE and BWS-2



Relative importance
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
dce_lc1
dce_lc2
dce_lc3
muscle strength
energy endurance
balance
cognition
blurry vision
liver damage

a.



b.

Figure 3 – Relative importance of attributes for DCE (a) and BWS-2 (b)

Table 3 – Latent class analysis results for DCE and BWS-2

n=140	DCE Class 1		DCE Class 2		DCE Class 3		BWS-2 Class 1		BWS-2 Class 2		BWS-2 Class 3	
	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se	β est.	Rob. Se
Muscle strength												
Stays the same	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
Improved by half	0.79	0.60	-0.40	0.51	0.98**	0.27	2.25**	0.50	3.48**	0.74	2.15**	0.42
Cured	0.94**	0.38	0.54	0.46	1.80**	0.35	3.01**	0.52	4.83**	0.88	3.00**	0.46
Energy endurance												
Stays the same	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
Improved by half	0.62**	0.27	0.96	0.70	0.08	0.26	2.67**	0.47	1.85**	0.66	2.83**	0.38
Cured	0.87*	0.48	1.63	1.58	0.77**	0.25	3.28**	0.44	3.19**	0.45	3.63**	0.43
Balance												
Stays the same	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
Improved by half	1.02*	0.52	-0.31	0.41	0.46	0.38	1.46**	0.41	2.52**	0.72	1.90**	0.35
Cured	1.40*	0.68	-0.43	0.46	1.21**	0.41	2.14**	0.49	3.65**	1.13	2.62**	0.40
Cognition												
Stays the same	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
Improved by half	0.10	0.28	0.69*	0.39	0.05	0.20	1.24**	0.44	1.32**	0.54	0.28	0.34
Cured	0.27	0.19	1.43	1.17	0.35	0.23	1.57**	0.47	2.49**	0.85	0.51	0.39
Chance blurry vision (%)												
1	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
15	-0.09	0.15	-0.18	0.39	-0.17	0.16	-0.49**	0.24	-0.16	0.13	-1.39**	0.34
30	-0.27*	0.15	-0.47	0.37	-0.56**	0.24	-1.32**	0.33	-0.97**	0.39	-2.27**	0.43
Chance liver damage (%)												
1	REF	-	REF	-	REF	-	REF	-	REF	-	REF	-
15	-0.51**	0.23	0.11	0.44	-3.12**	0.43	-1.81**	0.48	-0.33	0.24	-3.00**	0.44
30	-0.80*	0.43	-1.03	1.09	-6.56**	0.86	-2.45**	0.48	-1.43**	0.51	-4.30**	0.72
Reference levels[i]												
Muscle strength: stays the same	0	-	0	-	0	-	0	-	0	-	0	-
Energy endurance: stays the same	0	-	0	-	0	-	-0.21	0.30	0.69*	0.38	-1.13**	0.41
Balance: stays the same	0	-	0	-	0	-	0.13	0.30	0.58	0.51	-1.00**	0.44
Cognition: stays the same	0	-	0	-	0	-	0.64*	0.34	1.39**	0.57	-0.20	0.37
Chance blurry vision (%): 1	0	-	0	-	0	-	1.19**	0.42	4.40**	0.86	-0.65	0.58
Chance liver damage (%): 1	0	-	0	-	0	-	1.90**	0.41	5.60**	0.82	-2.05**	0.71
class shares	0.35		0.18		0.47		0.42		0.17		0.41	
delta_class[ii]	REF		-0.66		0.29		REF		-0.93		-0.03	
beta_worst[iii]	-						0.79					
log likelihood	-876.46						-5788.99					

* Significant at 10%
** Significant at 5%
[i] β estimates attribute levels estimated as additional utility or disutility compared to reference level
[ii] delta_class parameters indicate likelihood of being in specific class compared to reference class
[iii] beta_worst parameter allows for scale differences between best and worst choices (hypothesis testing beta_worst = 1 showed no statistically significant outcomes)

To characterize patients in the three different DCE and BWS-2 latent classes, ex-post analyses were conducted (Table 4) based on the descriptives in Table 2 since extending our LC model with a class membership model failed to converge due to the relatively small sample. These results show that DCE latent classes differed in terms of level of highest education, sex and age: DCE latent class 2 included the highest percentage females (72%), who were the youngest (median age 47) with the highest level of education (96% completed vocational or higher education). For BWS-2, latent class 2 was also different compared to other classes: this class included the highest percentage females (74%), who were the oldest (median age 58) and who were relatively less impaired by their disease (74% indicated to walk without an assistive device). The ex-post analyses in Table 4 also highlighted that there was a high level of concordance between patients in a specific DCE class and patients in the same BWS-2 class. More specifically, patients in the DCE class in which balance was the most important attribute (class 1) and in which liver damage was the most important attribute (class 3), had the highest probability to also be in BWS-2 latent class 1 (energy endurance most important) and latent class 3 (liver damage most important). This was however not the case for latent class 2.

Table 5 presents the results from the evaluation questions regarding DCE and BWS-2. The results show that there are no statistically significant differences between methods for evaluation questions about help with the survey, difficulty of answering questions and if the descriptions were sufficient. However, statistically significant (chi-squared test, p-value 0.04 < 0.05) differences were found between DCE and BWS-2 about difficulty of understanding the questions. A larger percentage of patients found it easier to understand DCE (74%) than BWS-2 (62%) questions.

Table 4 – Ex-post analyses of latent class analysis DCE and BWS-2

	DCE			BWS-2			DCE & BWS-2
Characteristic	**Class 1**	**Class 2**	**Class 3**	**Class 1**	**Class 2**	**Class 3**	**Overall sample**
Class share	35%	18%	47%	42%	17%	41%	-
Sex							
Female	65%	72%	63%	60%	74%	67%	65%
Age (years)							
Median (range)	54 (32-71)	47 (23-73)	55 (31-76)	54 (23-76)	58 (32-72)	53 (31-73)	54 (23-76)
Highest level of education							
No formal schooling	2%	0%	0%	0%	4%	0%	1%
Elementary	2%	0%	6%	3%	4%	3%	4%
Secondary	15%	4%	19%	21%	22%	7%	15%
Vocational	33%	40%	33%	29%	48%	34%	34%
Higher	44%	56%	42%	47%	22%	54%	45%
No answer	4%	0%	0%	0%	0%	2%	1%
Disease status							
Walk and run without assistive device	15%	24%	28%	19%	17%	30%	23%
Walk without assistive device	38%	40%	34%	32%	57%	31%	36%
Walk but rely on assistive device	33%	24%	21%	34%	9%	24%	26%
Walk but using wheelchair part-time	10%	12%	16%	12%	17%	14%	14%
Fully rely on wheelchair	4%	0%	1%	3%	0%	1%	1%
Crosstab	**BWS-2 class 1**	**BWS-2 class 2**	**BWS-2 class 3**				
DCE class 1	42%	21%	38%	-	-	-	-
DCE class 2	56%	20%	24%	-	-	-	-
DCE class 3	37%	12%	51%	-	-	-	-

Table 5 – Evaluation questions DCE and BWS-2

Evaluation question	DCE (n=131)[i]	BWS-2 (n=130)[i]
Help with survey (p-value = 1.00)		
By myself	120 (92%)	119 (92%)
Some help	10 (7%)	10 (7%)
Someone else	1 (1%)	1 (1%)
Difficulty understanding questions (p-value = 0.04)		
Very easy	41 (31%)	27 (21%)
Easy	56 (43%)	53 (41%)
Not easy or difficult	28 (21%)	34 (26%)
Difficult	6 (5%)	11 (8%)
Very difficult	0 (0%)	5 (4%)
Difficulty answering questions (p-value = 0.86)		
Very easy	19 (15%)	18 (13%)
Easy	48 (37%)	41 (32%)
Not easy or difficult	40 (31%)	41 (32%)
Difficult	22 (16%)	28 (22%)
Very difficult	2 (1%)	2 (1%)
Description of benefits and risks was sufficient (p-value = 0.41)		
Yes	119 (91%)	114 (88%)
No	12 (9%)	16 (12%)

[i] Difference in total number of patients in DCE and BWS-2 who completed the evaluation questions since these questions were no mandatory questions in the survey

9.4 Discussion

In this study preference weights and other outcomes (e.g., RIS) between DCE and BWS-2 were compared within NMD patients. We conclude that both methods lead to different preference weights as well as RIS values. However, accounting for preference heterogeneity, LC outcomes showed that patient classes look more similar, with a clear class of patients who both in DCE and BWS-2 indicated that liver damage was the most important attribute (class 3). For both preference methods, this class was among the largest classes of patients. Additionally, patients that identified liver damage as most important (class 3) in DCE also had the highest probability to be in the same class in BWS-2. The ex-post analyses also showed that for both preference methods class 2 differed in terms of descriptives (i.e., sex, age, education, disease status) compared to class 1 and class 3. Contrary to initial expectations, most patients found is easier to understand DCE than BWS-2 choice tasks.

One of our main findings of this study was that both DCE and BWS-2 led to different outcomes. There are several stated preference studies comparing outcomes between these two methods. Studies by Van Dijk et al. [11], Potoglou et al.[17] and Severin et al.[18] showed similar outcomes between DCE and BWS-2. Differences between these studies and our study might firstly be related to differences in the health decision context. Working with different type of respondents and dealing with different type of decisions (e.g., treatment choice, priority setting) might lead to different behaviour, different choices and therefore different outcomes. Secondly, in our study we explicitly framed negative attributes (i.e., blurry vision and liver damage) positively in BWS-2 choice tasks in order to avoid comparisons of positive and negative attributes with a BWS-2 choice task since this could lead to identification problems.[25] This was not the case in the previous studies. Additionally, there might also have been a framing effect in our study with regard to the attribute liver damage, since the word “permanent” was included in the choice task which might be a reason why this attribute was being considered important in both DCE and BWS-2. For the other negative attribute in DCE, risk of blurry vision, it was stated that problems would disappear once (hypothetical) medication would be stopped. Indeed, this temporary negative side-effect appeared to be far less important in patient decision-making. On the other hand, although our study differs from some of prior research studies comparing the two methods, our study outcomes are in line with a study by Whitty et al. [34] in which the authors also reported differences in relative preference weigths and preference orderings between DCE and BWS-2 in a priority setting context.

In our study the same patient sample (n=140) completed both 12 DCE and 18 BWS-2 choice tasks. Preference weights from LC in Table 3 showed that especially in DCE latent class 2 most attribute levels were not statistically significant (i.e., smaller t-values) compared to BWS-2. More in general, attribute levels in DCE overall had smaller t-values compared to BWS-2. This can be an indication that given the same (small) sample size, BWS-2 might be the preferred method of choice when statistical power is important for decision-making. It should be noted here that this can however only be conclued by assuming that the cognitive burden of the 12 DCE and 18 BWS-2 choice tasks are comparable. Our results also suggest a smaller utility scale for DCE, which suggests the need for larger sample sizes in DCE compared to BWS-2, as also mentioned in previous work.[24]

BWS-2 literature states that one of the reasons BWS-2 could be an interesting preference method compared to DCE is because of its lower cognitive burden.[11,12] However, this study indicated that patients found it easier to understand DCE compared to BWS-2 choice tasks. It should be noted here that the number of choice tasks between DCE (12) and BWS-2 (18) was different, which may have influenced the evaluation of the methods by patients. The findings in this study follow the trend as described in a study by Himmler et al.[35] in which the authors found that DCE choice tasks were less cognitively burdensome than

BWS-2 choice tasks. Whitty et al.[12] also reported that in their study the majority of respondents found it more difficult to complete BWS-2 compared to DCE and most respondents preferred DCE over BWS-2.

A strength of this study is that it is the first study focusing on differences in outcomes between DCE and BWS-2 with regard to a sample possibly hampered by cognitive limitations. As previously mentioned, several studies are focusing on differences between DCE and BWS-2 outcomes. However, to our knowledge, there are no such studies conducted within the context of a sample with cognitive limitations specifically. This study is also important, because NMD are considered rare diseases which often translates into relatively small sample sizes when eliciting preferences. This study provides useful insights into how BWS-2 and DCE performed with a relatively small sample size.

At the same time, the relatively small sample size is a limitation of this study. In general, this will not be a problem when estimating choice models not accounting for preference heterogeneity (MNL). However, when estimating more sophisticated models like for example LC in this study, such small sample sizes could potentially lead to estimation problems. In this study we were able to estimate an LC model, but the extension with a class membership model failed to converge. Therefore, descriptive ex-post analyses were conducted to characterize the different latent patient classes. Future studies should however focus on larger samples that have cognitive limitations to investigate preference heterogeneity more thoroughly.

9.5 Conclusion

This study showed that using either DCE or BWS-2 leads to different preference weights as well as relative importance values. A potential reason lies in the way risks were framed (i.e., positive) in BWS-2, which was different than in DCE. Patients indicated that DCE choice tasks were not more difficult to understand than BWS-2 tasks. However, accounting for preference heterogeneity, the latent class analysis indicated latent classes in both DCE and BWS-2 that are comparable, especially the class of patients that indicated that liver damage was the most important attribute. Hence, we advise careful consideration when selecting either BWS-2 or DCE to elicit preferences since our results suggest that BWS-2 is the preferred method of choice when dealing with small samples, while DCE may be preferred when minimizing cognitive burden is key and choice tasks include both benefits and risks. It will therefore be important that the method matches the decision context. To support medical decision-making, keeping in mind the research and decision context will be key.

References

1. Hoos A, Anderson J, Boutin M, et al. Partnering With Patients in the Development and Lifecycle of Medicines: A Call for Action. *Ther Innov Regul Sci*. 2015;49(6):929-939. doi:10.1177/2168479015580384
2. Anderson RM, Funnell MM. Patient empowerment: reflections on the challenge of fostering the adoption of a new paradigm. *Patient Educ Couns*. 2005;57(2):153-157. doi:10.1016/J.PEC.2004.05.008
3. Smith MY, Hammad TA, Metcalf M, et al. Patient Engagement at a Tipping Point-The Need for Cultural Change Across Patient, Sponsor, and Regulator Stakeholders: Insights From the DIA Conference, "Patient Engagement in Benefit Risk Assessment Throughout the Life Cycle of Medical Products". *Ther Innov Regul Sci*. 2016;50(5):546-553. doi:10.1177/2168479016662902
4. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. *Patient - Patient-Centered Outcomes Res*. 2017;10(3):263-266. doi:10.1007/s40271-017-0222-3
5. Food and Drug Administration (FDA). The Voice of the Patient: A Series of Reports from FDA's Patient-Focused Drug Development Initiative | FDA. Published 2017. Accessed June 8, 2021. https://www.fda.gov/industry/prescription-drug-user-fee-amendments/voice-patient-series-reports-fdas-patient-focused-drug-development-initiative
6. (MDIC) MDIC. MEDICAL DEVICE INNOVATION CONSORTIUM (MDIC) PATIENT CENTERED BENEFIT-RISK PROJECT REPORT: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology By Medical Device I. Accessed July 8, 2018. http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf
7. Soekhai V, Whichello C, Levitan B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324-1331. doi:10.1016/j.drudis.2019.05.001
8. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-189. doi:10.1016/j.jhealeco.2006.04.002
9. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):1-14. doi:10.1186/s13561-015-0079-x
10. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*.; 2015. doi:10.1017/CBO9781107337855
11. van Dijk JD, Groothuis-Oudshoorn CGM, Marshall DA, IJzerman MJ. An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. *Value Heal*. 2016;19(4):316-322. doi:10.1016/j.jval.2015.12.020
12. 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. *Med Decis Mak*. 2014;34(5):638-654. doi:10.1177/0272989X14526640
13. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics*. 2014;32(9):883-902. doi:10.1007/s40273-014-0170-x
14. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics*. 2019;37(2):201-226. doi:10.1007/s40273-018-0734-2
15. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*. Cambridge University Press; 2015. doi:10.1007/9781316136232
16. Train K. *Discrete Choice Methods with Simulation*.; 2002.
17. 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(10):1717-1727. doi:10.1016/j.socscimed.2011.03.027
18. 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. *Eur J Hum Genet*. 2013;21(11):1202-1208. doi:10.1038/ejhg.2013.36
19. Weber YG, Roebling R, Kassubek J, et al. Comparative analysis of brain structure, metabolism, and cognition in myotonic dystrophy 1 and 2. *Neurology*. 2010;74(14):1108-1117. doi:10.1212/WNL.0b013e3181d8c35f
20. Lax NZ, Gorman GS, Turnbull DM. Review: Central nervous system involvement in mitochondrial disease. *Neuropathol Appl Neurobiol*. 2017;43(2):102-118. doi:10.1111/nan.12333
21. Johnson N, Imbrugia C, Dunn D, et al. Genetic Prevalence of Myotonic Dystrophy Type 1 (S23.003).

Neurology. 2019;92(15 Supplement).
22. Gorman GS, Schaefer AM, Ng Y, et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol*. 2015;77(5):753-759. doi:10.1002/ana.24362
23. Meola G, Sansone V. Cerebral involvement in myotonic dystrophies. *Muscle and Nerve*. 2007;36(3):294-306. doi:10.1002/mus.20800
24. 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. *J Choice Model*. 2013;6:34-43. doi:10.1016/j.jocm.2013.04.004
25. Soekhai V, Donkers B, de Bekker-Grob E. PNS295 BEST WORST SCALING: FOR GOOD OR FOR BAD BUT NOT FOR BOTH. *Value Heal*. 2019;22:S813. doi:10.1016/j.jval.2019.09.2195
26. Cecilia Jimenez-Moreno A, Anne Pinto C, Levitan B, et al. A study protocol for quantifying patient preferences in neuromuscular disorders: a case study of the IMI PREFER Project [version 1; peer review: 1 approved]. Published online 2020. doi:10.12688/wellcomeopenres.16116.1
27. Jimenez-Moreno AC, van Overbeeke E, Pinto CA, et al. Patient Preferences in Rare Diseases: A Qualitative Study in Neuromuscular Disorders to Inform a Quantitative Preference Study. *Patient - Patient-Centered Outcomes Res*. Published online February 27, 2021:1-12. doi:10.1007/s40271-020-00482-z
28. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Heal*. 2013;16(1):3-13. doi:10.1016/j.jval.2012.08.2223
29. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*.
30. De Bekker-Grob EW, Hol L, Donkers B, et al. Labeled versus unlabeled discrete choice experiments in health economics: An application to colorectal cancer screening. *Value Heal*. 2010;13(2):315-323. doi:10.1111/j.1524-4733.2009.00670.x
31. Hahn G, Shapiro S. *A Catalogue and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Designs*. General Electric Research and Development Center; 1966.
32. Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *J Choice Model*. 2019;32:1-43. doi:10.1016/j.jocm.2019.100170
33. Hess S, Palma D. Apollo : a flexible , powerful and customisable freeware package for choice model estimation and application Stephane Hess & David Palma Choice Modelling Centre University of Leeds. *Apollo user Man*. 2019;(April).
34. 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(4):e90635. doi:10.1371/journal.pone.0090635
35. Himmler S, Soekhai V, van Exel J, Brouwer W. What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. *J Choice Model*. 2021;38:100265. doi:10.1016/j.jocm.2020.100265

Chapter 10

General discussion

10 Choice modelling in health: discussion

This dissertation explores and addresses several methodological challenges and opportunities for choice modelling – specifically DCE and BWS-2 – in health. First, the main findings in relation to the objectives of this dissertation will be presented. Second, several points of interest for choice modelling in health will be discussed. This chapter will end with conclusions and recommendations for future research.

10.1 Main findings

Objective 1a: Providing an overview of current preference methods used in health
Objective 1b: Providing an overview of DCE applications in health

Chapter 2 provided an up-to-date compendium and taxonomy of both preference exploration (qualitative) and elicitation (quantitative) methods within the medical product lifecycle (MPLC) by making use of a three-step approach to identify existing preference methods. In total, 32 unique preference methods were identified and grouped into several categories. The developed compendium and taxonomy can serve as an important resource for assessing these methods and helping to determine which are most appropriate for different research questions. The findings from this study are partly in line with prior studies aiming to provide an overview of preference exploration and elicitation methods.[1,2] Differences are mostly due to the focus of the review (i.e., patient preferences only or broader), focus on preference elicitation methods only and methodology used to identify methods.

In **chapter 3** an overview and description of trends, current practice and future challenges of the applications and methods used by discrete choice experiments (DCE) in health economics was given (period 2013-2017). In total, 301 DCE studies were published covering a range of policy questions: valuing patient experience, valuing health outcomes, trade-offs between patient experience and health outcomes, estimating utilities within QALY context, priority setting and preferences regarding clinical decision-making. This means that not only the total number of DCEs per year continued to increase, but the application also broadened with an increased geographic scope. In **chapter 3** we also reported that more sophisticated experimental designs and specific software to generate these designs were used. The trend towards using more sophisticated econometric models also continued. We believe this is a positive development, since more sophisticated techniques are generally better able to approximate true preferences. For example when preference heterogeneity is assumed and MXL is used for modelling instead of MNL. However, many studies presented sophisticated methods with insufficient detail, making it hard to reproduce and check study outcomes.

Objective 2a: Providing insights whether choice share predictions in DCE depend on modelling and analysis approach

In **chapter 4** the accuracy of DCE choice share predictions was studied. More specifically, we studied whether these predictions depended on the econometric modelling approach used and type of analysis being conducted, while dealing with preference heterogeneity. Results from our simulation study indicated that models that did and did not account for preference heterogeneity can be used both to obtain estimates and predict choice shares. However, relying on a model that did not account for preference heterogeneity at all (multinomial logit, MNL) performs better (i.e., lower error in predicted choice shares) compared to a more complicated model that did account for preference heterogeneity (mixed logit, MXL) when heterogeneity is ignored in the prediction stages. The accuracy of DCE choice share predictions is maximized when using both MXL and accounting for preference heterogeneity in the prediction of choice shares. Our outcomes showed that conducting non-corresponding analyses after model estimation hugely impacts outcomes. This means additional skills from researchers or added functionality of available software packages is needed to make sure preference heterogeneity is also accounted for in post-estimation analyses (assuming preference heterogeneity is relevant and should be accounted for).

Objective 2b: Providing insights into the impact of mixing positive and negative attributes in BWS-2
Objective 2c: Providing insights into the impact of framing in BWS-2

We showed in **chapter 5** that mixing positive and negative attributes (e.g., studying treatments characterized by both benefits and harms) in case 2 best-worst scaling (BWS-2) intuitively leads to attribute dominance. It was analytically shown that dominance leads to infinitely large differences between the parameter estimates for the positive versus negative attributes and therefore estimation problems. Our simulation results confirmed our analytical findings and showed that parameter values could not be accurately recovered and also led to problems with the relative ordering of attribute levels. To potentially overcome these issues, we provided a solution in **chapter 6**. In this study attributes were either framed all positive, all negative or mixed. Our results showed differences in outcomes between positively and negatively framed BWS-2, with differences in the preference pattern for the least preferred attribute level. Differences in the ordering of attribute levels between the framings was also found. Results also showed that framing attributes negatively led to attributes becoming more important, while framing attributes positively led to attributes becoming less important. This study also provided evidence that mixing positive and negative attributes in BWS-2 leads to theoretically implausible outcomes (e.g., 85% being more important than 100% chance of being cured).

Objective 2d: Providing insights into including explicit reference points in BWS-2

In **chapter 7** we aimed to introduce a new BWS-2 approach by including explicit reference points in the choice tasks (BWS-2R) and to investigate whether this new approach led to a more accurate analysis of preferences compared to BWS-2. We showed analytically that BWS-2R should reduce noise in the inferred preferences as regular BWS-2 preferences are confounded with reference points. Our study results showed statistically significant differences between estimated preference weights for both BWS-2 approaches, as well as statistically significant differences in relative importance scores (RIS) between BWS-2R and BWS-2. No difference in perceived difficulty between the two approaches was found, although a larger proportion of respondents that completed BWS-2 preferring BWS-2R than the other way around.

Objective 3a: Providing insights into differences in perceived cognitive burden between DCE and BWS-2

Objective 3b: Providing insights into differences in statistical outcomes between DCE and BWS-2

Results from a randomized experiment among older people from the general population showed that using different visual presentations of DCE and BWS-2 choice tasks (i.e., level-overlap and color-coded), DCE tasks were found to be less cognitive burdensome compared to BWS-2 (**chapter 8**). Especially color-coding in BWS-2 could not be recommended because its effect on cognitive burden was not clear and led to undesired choice heuristics. In **chapter 9** DCE and BWS-2 were empirically compared in terms of statistical outcomes.

Results from this study indicated differences in preference weights as well as RIS attribute rankings between DCE and BWS-2. On the other hand, latent class analysis revealed similar latent classes for the same sample of patients, but patients did not find the DCE more difficult than BWS-2. Our results suggest that BWS-2 is the preferred method of choice when dealing with small samples, while DCE may be preferred when minimizing cognitive burden is key and choice tasks include both benefits and risks.

10.2 Points of interest for choice modelling in health

Similar to other fields, applying methods from choice modelling in health provides benefits but also comes with challenges. In this section several points of interest regarding the studies that are part of this dissertation will be addressed and discussed.

10.2.1 Comparability of studies

The two studies (**chapters 2-3)** in which a systematic literature review was the core method to collect data, might suffer some drawbacks. In the interests of a time-efficient and precise review, synonyms for the systematic literature review were limited since prior reviews were also analyzed and international experts were consulted (**chapter 2**). To also ensure comparability between previous DCE reviews (2001-2012), searches to identify DCEs in health economics between 2013 and 2017 were restricted to PubMed only (**chapter 3**).[3–5]

10.2.2 Reporting of methodological details in DCE

Using more sophisticated techniques that provide a way to for example account for preference heterogeneity provides outcomes that more closely represent true preferences.[6,7] More and more DCE applications in health make use of more sophisticated techniques (**chapter 3**). The presence of such studies suggests that the knowledge of DCEs in health is increasing, leading to the use of more sophisticated techniques for experimental design generation and econometric modelling. Given this trend, it is crucial for DCE researchers to provide sufficient detail regarding experimental design choices (e.g., providing information about the design plan) and econometric modelling (e.g., the amount of Halton draws when using mixed logit (MXL) for estimation). Omitting this type of information might inhibit quality assessment, reduce confidence in the results and therefore reduces the ability of decision-makers to act on the results. Providing this kind of information also helps the field of choice modelling in health to become more mature.

10.2.3 Post-estimation analyses in DCE

It is also important to provide insights into the implications of the statistical analysis of choice data, especially for more policy relevant measures like predicted choice shares and accounting for preference heterogeneity in the post-estimation analysis phase. Preference heterogeneity reflects differences in preferences within a population and the importance of preference heterogeneity on the application of patient preferences has been widely discussed in the literature.[6,8,9]

As demonstrated in **chapter 4**, the accuracy of choice share predictions in DCE depends on the modelling approach (multinomial logit (MNL) or MXL) and the type of post-estimation analysis (accounting for preference heterogeneity or not). Where the MNL model assumes that individuals have identical preferences, the MXL assumes a distribution of preferences with individual specific preferences.[6,7,10] When executed correctly, choice shares predictions based on MXL are overall more accurate than those of MNL.[11,12] "Executing correctly" in this context includes also accounting for preference heterogeneity after estimation. This means that after using MXL to estimate coefficients, not the estimated mean preference parameters should be used for choice share predictions. Instead, the full distribution of individual-level preferences should be used for predictions. That way the inferred heterogeneity in preferences is not ignored in the post-estimation stage. After all, using MXL instead of MNL to estimate preference parameters demands larger sample sizes and computation times.[13,14] It would therefore make no sense to go through all this extra effort in the estimation stage, but not in the post-estimation stage when computing relevant measures.

It is important to provide more insights into potential problems that could arise in the post-estimation stage, since multiple studies focused on the estimated population mean preference parameters to predict choice shares in order to perform policy simulations and recommendations.[15,16] Therefore, we argue that additional researcher knowledge and skills are needed – or additional added functionality of available software packages – to correctly predict choice shares.

10.2.4 Framing attributes in BWS-2

Case 2 best-worst scaling (BWS-2) has become more often used in health, although the method itself is still in its infancy and several issues related to the design and analyses require further research[17–1920] One of these issues is related to the framing of attribute levels in BWS-2 as a potential solution to overcome issues related to attribute dominance as a result of mixing positive and negative attributes in BWS-2 choice tasks (**chapter 5**). Avoiding attribute dominance by for example focusing on attribute framing in BWS-2 is especially important since this is not a trade-off method like a DCE.[6,21,22] Framing in this context explicitly refers to an all positive or all negative framing of the BWS-2 attribute levels. In a DCE dominance can for example be avoided as long as the attractiveness of the best attribute level can be compensated by the inclusion of attractive levels of the other attributes in the other choice alternatives. In BWS-2, the choices are at the attribute level, not at the profile level, hence such compensation is not possible. As soon as a single attribute level is strictly preferred over all others, dominance cannot be avoided, as there is no possibility to compensate through a combination with less attractive levels for the other attributes.

Framing attributes could however get complicated as it is well-known from behavioral economic theory that individuals cope differently when dealing with (hypothetical) gains (e.g. increased life expectancy) or losses (e.g. more frequent side effects) and we know from previous studies focused on attribute framing that this impacts preference research outcomes[17–19,23] In general, individuals tend to place more weight on losses than similar sized gains. The results from our study in **chapter 6** are in line with theory-based predictions: attributes become less important when framed positively and more important when framed negatively. This means that the type of framing will impact BWS-2 estimation outcomes and therefore conclusions. Hence, we recommend to carefully consider how to frame attributes in BWS-2 experiments. It will be important that the frame matches the decision environment that the respondent is facing as well as the fact that attributes and levels should be always presented in such a way that they are relevant and relatively easy to interpret. Qualitative work and pretesting will be important to validate this.

Gaining insights into how to evaluate the impact of a frame on preferences in health has been studied before in psychological literature, by presenting individuals with both the positive and negative frame.[24,25] More detailed information about the impact of framing on preferences will increase the understanding of and confidence in preference outcomes, which will improve our understanding of framing effects in medical decision-making.[26]

10.2.5 A new BWS-2 approach

As mentioned in the previous section, there are several issues related to BWS-2 that require further exposition before these outcomes might get a more prominent role in health decision-making. One of the issues we outlined in **chapter 7** is the undesirable role of differences in reference points that drive BWS-2 outcomes. BWS-2 is used to elicit preferences and we want that differences in preferences represent actual differences in preferences and not driven by (unobserved) reference point differences. We therefore introduced a new BWS-2 approach with explicit reference points (BWS-2R) to reduce the noise in the inferred preferences as BWS-2R preferences are no longer confounded with differences in reference points. Studies by Stathopoulos & Hess and Hess et al.[27,28] previously showed that reference points impact preferences of individuals.

The interpretation of BWS-2R outcomes becomes different from BWS-2: BWS-2R outcomes should be interpreted keeping the explicit reference point in mind. This means a full ranking of all attribute levels is not possible due to the relationship between attribute level and specific reference point in BWS-2R, with attribute levels no longer being directly comparable (which is possible in BWS-2). On the other hand, differences between attribute levels can still be interpreted and we believe the problems associated with completely unknown reference points in BWS-2 will be much larger than the more complex interpretation BWS-2R demands. Testing whether the external validity (whether individuals behave in reality as they state in a hypothetical choice experiment) of BWS-2 actually improves when the BWS-2R approach is used will also be important for practical use.

10.2.6 Perceived cognitive burden of choice modelling in health

We noticed differences in perceived cognitive burden between DCE and BWS-2 as well as differences in statistical outcomes (i.e., preference weights and relative importance scores) between the methods (**chapters 8-9**). The relevant question of “which method is better to elicit preferences” is not an easy one to answer. This really depends on the context and available resources. We could for example imagine that BWS-2 is the preferred method of choice when dealing with relatively small sample sizes (our results suggest that DCEs demand larger samples), while DCE is preferred when positive and negative attributes are involved and attribute framing is troublesome. There is also evidence that individuals find it more difficult to complete BWS-2 compared to DCE (**chapter 9** and study by Whitty et al.).[29] Both choice modelling methods can however provide valuable insights into individual's health preferences, which can be used in actual health decision-making.

There are however two major factors that complicate the use of choice modelling in health. First, using either DCE or BWS-2 requests knowledge of experimental design theory and econometric modelling. This could easily become complex. Although methodological knowledge of DCEs in health is likely to increase, driven by the increasing number of DCE applications and more and more methodological DCE studies in health becoming available (**chapter 3**), we are convinced there are still knowledge gaps how to conduct choice experiments in a scientific sound way and to critically assess outcomes. These methods also require quite a lot of preparation and are usually combined with a qualitative phase to for example identify attributes and levels.

In the DCE studies included in the systematic literature review (**chapter 3**), only a handful of studies reported information about the external validity. Although the external validity of DCEs has been studied in different health contexts, similar studies for BWS-2 are lacking.[30–34] For DCEs in health, external validity results are promising with respect to the predictions on an aggregate level.[35] Although steps in the right direction are taken, with more methodological DCE and BWS-2 studies in the health context, external validity of choice modelling in health remains a under-researched topic. One of the reasons that has been stated before is the lack of available real choice (revealed-preference) data.[36] With the advent of approaches that could reduce the hypothetical bias (hypothetical behaviour does not necessarily match actual behaviour) from stated-preference outcomes by using revealed preference data that informs the experimental design and econometric modelling, the overall external validity of choice modelling in health can be studied in more depth.[37] This is in line with trends of more studies about the role of using (real) patient journey data for decision-making in health.[38] This could potentially increase the external validity of choice modelling outcomes, which would improve the confidence of decion-makers in these outcomes and adds to maturing the field of choice modelling in health.

10.3 Conclusions and recommendations for future research

Conclusions

- Preference exploration (qualitative) and elicitation (quantitative) methods in health can be divided into several categories from which DCE and BWS-2 are popular methods
- The total number of DCEs in health per year continued to increase, with a broadened area of application, increased geographic scope and the use of more sophisticated experimental designs and econometric modelling.
- When dealing with preference heterogeneity and predicting choice shares in DCE, preference heterogeneity should also be taken into account in the prediction stage
- Mixing positive and negative attributes in BWS-2 should be avoided since this leads to attribute dominance and therefore estimation problems
- A potential solution to avoid attribute dominance due to mixing positive and negative attributes in BWS-2 is to frame attributes either all positive or all negative
- Using BWS-2R instead of BWS-2 should reduce noise in the inferred preferences as BWS-2 preferences are confounded with reference points
- DCE was found to be less cognitive burdensome compared to BWS-2 in an experiment among older people

- Although, there were differences in statistical outcomes between DCE and BWS-2, latent classes of preference patterns were similar to each other.

Future research is needed:

- To determine how and when to integrate choice modelling outcomes into health decision-making
- To provide insights into how DCE specific reporting guidelines could contribute to quality assessment of DCE results, increase confidence in the results and improve the ability of decision-makers to act on the results.
- To further provide evidence for the loss in accuracy of predicted choice shares in DCE when not taking preference heterogeneity into account in the prediction stage
- To improve the understanding and use of attribute framing in BWS-2 to avoid estimation problems
- To find new possibilities to test the external validity of choice modelling outcomes in health, especially with new approaches like BWS-2R

References

1. Ryan, M R, Scott, DA S, Reeves, C R, et al. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*. 2001;5(5). doi:10.3310/HTA5050

2. (MDIC) MDIC. MEDICAL DEVICE INNOVATION CONSORTIUM (MDIC) PATIENT CENTERED BENEFIT-RISK PROJECT REPORT: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology By Medical Device I. Accessed July 8, 2018. http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf

3. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64. Accessed June 18, 2018. http://www.ncbi.nlm.nih.gov/pubmed/14619274

4. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145-172. doi:10.1002/hec.1697

5. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics*. 2014;32(9):883-902. doi:10.1007/s40273-014-0170-x

6. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*. Cambridge University Press; 2015. doi:10.1007/9781316136232

7. Train K. *Discrete Choice Methods with Simulation*. Cambridge University Press; 2002.

8. Hess S, Meads D, Twiddy M, Mason S, Czoski-Murray C, Minton J. Characterising heterogeneity and the role of attitudes in patient preferences: A case study in preferences for outpatient parenteral intravenous antimicrobial therapy (OPAT) services. *J Choice Model*. 2021;38:100252. doi:10.1016/j.jocm.2020.100252

9. Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *J Health Econ*. 2008;27(4):1078-1094. doi:10.1016/j.jhealeco.2007.11.006

10. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Heal*. 2016;19(4):300-315. doi:10.1016/j.jval.2016.04.004

11. Hess S, Polak JW. Mixed logit modelling of airport choice in multi-airport regions. *J Air Transp Manag*. 2005;11(2):59-68. doi:10.1016/j.jairtraman.2004.09.001

12. Alfnes F. Stated preferences for imported and hormone-treated beef: Application of a mixed logit model. *Eur Rev Agric Econ*. 2004;31(1):19-37. doi:10.1093/erae/31.1.19

13. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *Patient*. 2015;8(5):373-384. doi:10.1007/s40271-015-0118-z

14. Hess S, Train KE. Recovery of inter- and intra-personal heterogeneity using mixed logit models. *Transp Res Part B Methodol*. 2011;45(7):973-990. doi:10.1016/j.trb.2011.05.002

15. Determann D, Korfage IJ, Lambooij MS, et al. Acceptance of vaccinations in pandemic outbreaks: A discrete choice experiment. *PLoS One*. 2014;9(7):e102505. doi:10.1371/journal.pone.0102505

16. Grauman Å, Hansson · Mats, James S, Hauber · Brett, Veldwijk J. The Patient-Patient-Centered Outcomes Research Communicating Test Results from a General Health Check: Preferences from a Discrete Choice Experiment Survey. *Patient - Patient-Centered Outcomes Res*. Published online 123AD. doi:10.1007/s40271-021-00512-4

17. Tversky A, Kahneman D. *The Framing of Decisions and the Psychology of Choice*.; 1981.

18. Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value Heal*. 2009;12(2):354-363. doi:10.1111/j.1524-4733.2008.00417.x

19. Veldwijk J, Essers BAB, Lambooij MS, Dirksen CD, Smit HA, De Wit GA. Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? *Value Heal*. 2016;19(2):202-209. doi:10.1016/j.jval.2015.11.004

20. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):1-14. doi:10.1186/s13561-015-0079-x

21. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*.; 2015. doi:10.1017/CBO9781107337855

22. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):259-267. doi:10.1586/erp.10.29

23. Diederich A, Wyszynski M, Traub S. Need, frames, and time constraints in risky decision-making. *Theory Decis 2020 891*. 2020;89(1):1-37. doi:10.1007/S11238-020-09744-6

24. Kühberger A. The Framing of Decisions: A New Look at Old Problems. *Organ Behav Hum Decis Process*. 1995;62(2):230-240. doi:10.1006/OBHD.1995.1046

25. Druckman JN. Evaluating framing effects. *J Econ Psychol*. 2001;22(1):91-101. doi:10.1016/S0167-4870(00)00032-5

26. Gong J, Zhang Y, Yang Z, Huang Y, Feng J, Zhang W. The framing effect in medical decision-making: a review of the literature. *http://dx.doi.org/101080/135485062013766352*. 2013;18(6):645-653. doi:10.1080/13548506.2013.766352

27. Stathopoulos A, Hess S. Revisiting reference point formation, gains-losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. Transp Res Part A Policy Pract. 2012;46(10):1673-1689. doi:10.1016/j.tra.2012.08.005

28. Hess S, Rose JM, Hensher DA. Asymmetric preference formation in willingness to pay estimates in discrete choice models. Transp Res Part E Logist Transp Rev. 2008;44(5):847-863. doi:10.1016/j.tre.2007.06.002

29. 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. *Med Decis Mak*. 2014;34(5):638-654. doi:10.1177/0272989X14526640

30. Kesternich I, Heiss F, McFadden D, Winter J. Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D. *J Health Econ*. 2013;32(6):1313-1324. doi:10.1016/J.JHEALECO.2012.11.006

31. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in wales: A discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*. 2013;31:345–55.

32. Krucien N, Gafni A, Pelletier-Fleury N. Empirical Testing of the External Validity of a Discrete Choice Experiment to Determine Preferred Treatment Option: The Case of Sleep Apnea. *Health Econ*. 2015;24(8):951-965. doi:10.1002/HEC.3076

33. Quaife M, Terris-Prestholt F, Tanna GL Di, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *Eur J Heal Econ 2018 198*. 2018;19(8):1053-1066. doi:10.1007/S10198-018-0954-6

34. de Bekker-Grob EW, Swait JD, Kassahun HT, et al. Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. *Value Heal*. 2019;22(9):1050-1062. doi:10.1016/j.jval.2019.04.1924

35. de Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait JD. Can healthcare choice be predicted using stated preference data? *Soc Sci Med*. 2020;246:112736. doi:10.1016/J.SOCSCIMED.2019.112736

36. Lancsar E, Swait J. Reconceptualising the External Validity of Discrete Choice Experiments. *Pharmacoeconomics*. 2014;32(10):951-965. doi:10.1007/s40273-014-0181-7

37. Corte K de, Cairns J, Grieve R. Stated versus revealed preferences: An approach to reduce bias. *Health*

Econ. 2021;30(5):1095-1123. doi:10.1002/HEC.4246

38. Gualandi R, Masella C, Piredda M, Ercoli M, Tartaglini D. What does the patient have to say? Valuing the patient experience to improve the patient journey. *BMC Heal Serv Res 2021 211*. 2021;21(1):1-12. doi:10.1186/S12913-021-06341-3

Summary

The choice modelling field is concerned with understanding how individuals make choices by quantifying the underlying preferences. More specifically, choice modelling aims to characterize choices individuals make and to predict choices among the choice alternatives considered. Choice modelers assume that choices from individuals are based on preferences determining the amount of satisfaction (utility) they derive from goods and services. In choice modelling, individual's choices are related to their preferences by focusing on the utilities of choice alternatives. Choice modelling (i.e., specifically discrete choice experiments) was introduced in health in the early 1990s, especially to capture outcomes beyond health for health benefit assessments. Before the introduction of choice modelling in health, methods from other fields were used to gain insights into health preferences. Since choice modelling provided a new way to gain insights into health preferences the application of choice modelling in health continued to grow after the 1990s, especially with regard to stated-preferences.

There are several stated-preference methods to gain insights into health preferences, but discrete choice experiments (DCE) are increasingly advocated. A DCE is a survey-based preference elicitation method in which individuals are asked to select their preferred alternative from a set of hypothetical alternatives. DCE data analysis has its origin in mathematical psychology, with wide applications in marketing, transport and environmental economics and its theoretical foundation in random utility theory (RUT). Best-worst scaling (BWS) is another stated-preference method that has become an increasingly popular method to elicit health preferences. The introduction of BWS came from the intent to obtain more preference information than from a DCE by asking individuals to select their "best" and "worst" option, without increasing the cognitive burden. There are several types of BWS, but case 2 BWS (BWS-2) received much attention in the literature since this method can uncover attribute level importance, reduce cognitive burden of the choice task by focusing on one profile at a time and experiments are relatively easy to design. More detailed information about these methods can be found in **chapter 1** of this dissertation.

DCE and BWS-2 gained popularity in health to elicit preferences, although there are several methodological challenges to overcome inhibiting these methods to become more valuable for actual decision-making. Therefore, gaining insights into methodological challenges and providing opportunities to overcome them, contributes to academic literature as well as practice to inform decision-making. This dissertation has three main objectives (**chapter 1**):

1. Providing insights into preference methods used in health
2. Providing insights into DCE and BWS-2 challenges and opportunities regarding design and analysis
3. Empirically comparing outcomes between DCE and BWS-2

Insights into preference methods used in health

Chapter 2 provides an up-to-date compendium and taxonomy of both preference exploration (qualitative) and elicitation (quantitative) methods since it is important to further drive research on the incorporation of preferences in health decision-making forward. A three-step approach was used to identify existing preference methods. First, a systematic literature review of 4,572 unique papers identified through multiple scientific databases was conducted, using English full-text papers published between 1980 and 2016. Second, prior preference method reviews were examined to cross validate these results. Third, international experts (n=24) were consulted to confirm these results and to detect other potential methods. The systematic review (19 methods), analysis of prior conducted preference method reviews (23 methods), and expert consultations (4 methods) contributed to the compendium. In total, 32 unique methods were identified. These methods were grouped into several categories to provide an up-to-date compendium and taxonomy.

Chapter 3 gives an overview and description of trends, current practice and future challenges of the applications and methods used by DCEs in health economics since three previously published systematic literature reviews (covering time period 1990-2012). This review provides information whether the challenges identified in prior reviews are still relevant or whether there has been a response to the published suggestions and guidelines since key barriers to wider use of DCEs in policy included concerns about the robustness and validity of the method and the quality of applied studies. Between 2013-2017, 301 studies were identified that met the inclusion criteria and underwent data extraction. The results showed that the total number of DCEs per year continued to increase, with broader areas of application and increased geographic scope. Studies also reported using more sophisticated experimental designs and specific software to generate. The trend towards using more sophisticated econometric models also continued. However, many studies presented sophisticated methods with insufficient detail. Additionally, qualitative research methods continued to be a popular approach for identifying attributes and levels. However, inadequate reporting of methodological details could inhibit quality assessment and therefore may reduce decision-maker's confidence to act on DCE findings.

Insights into DCE and BWS-2 challenges and opportunities regarding design and analysis

Chapters 4-7 of the dissertation include DCE and BWS-2 studies focusing on methodological challenges and opportunities related to the design of DCE and BWS-2 experiments, as well as the analysis of DCE and BWS-2 choice data. In **chapter 4** the accuracy of DCE choice share predictions was studied: investigating whether these predictions depended on the econometric modelling approach used and type of analysis being conducted, while dealing with preference heterogeneity. Results from the simulation study indicated that, models that did and did not account for preference heterogeneity can be used to obtain estimates and predict choice shares. However, relying on a model that did not account for preference heterogeneity at all (multinomial logit, MNL) performs better (i.e., lower error in predicted choice shares) compared to a more complicated model that did account for preference heterogeneity (mixed logit, MXL) when heterogeneity is ignored in the prediction stages. The accuracy of DCE choice share predictions is maximized when using MXL and accounting for preference heterogeneity in the prediction of choice shares. These outcomes showed that conducting non-corresponding analyses hugely impacts outcomes and therefore decision-making.

In **chapter 5** the performance of BWS-2 when it is applied to a mix of positive and negative attributes (e.g., studying treatments characterized by both benefits and harms) is studied. This mix intuitively leads to attribute dominance. It was analytically showed that dominance leads to infinitely large differences between the parameter estimates for the positive versus negative attributes and therefore estimation problems. The simulation study confirmed our analytical results: parameter values of the attributes could not be accurately recovered. When only a single positive attribute was used, even the relative ordering of the attribute level preferences was not identified. **Chapter 6** provided a potential solution to the problem of mixing positive and negative attributes in BWS-2. In this study attributes were either framed all positive, all negative or mixed. A total of 192 patients were included for analysis, indicating differences in outcomes between positively and negatively framed BWS-2. Results also showed that framing attributes negatively led to attributes becoming more important, while framing attributes positively led to attributes becoming less important. This study also provided evidence that mixing positive and negative attributes in BWS-2 leads to theoretically implausible outcomes.

The final chapter of this section (**chapter 7**) presents a study which aimed to introduce a new BWS-2 approach including explicit reference points in the choice tasks (BWS-2R) and to investigate whether this new approach led to a more accurate analysis of preferences compared to BWS-2. It was analytically showed that BWS-2R should reduce noise in the inferred preferences as regular BWS-2 preferences are confounded with reference points. Our empirical study results (n=601) showed statistically significant differences between estimated preference weights for both BWS-2 approaches, as well as statistically significant differences in relative importance scores (RIS) between BWS-2R and BWS-2. No difference in perceived difficulty between the two approaches was found, with a larger proportion of respondents that completed BWS-2 preferring BWS-2R than the other way around.

Empirically comparing DCE and BWS-2 outcomes

Chapter 8 presents an empirical study comparing the perceived cognitive burden between DCE and BWS-2. Results from a randomized experiment (n=469) showed that using different visual presentations of DCE and BWS-2 choice tasks (i.e., level-overlap and color-coded), DCE tasks were found to be less cognitive burdensome compared to BWS-2. Especially color-coding in BWS-2 could not be recommended because its effect on cognitive burden was not clear and led to undesired choice heuristics. In **chapter 9** DCE and BWS-2 were empirically compared in terms of statistical outcomes. Results from this empirical study (n=140) indicated differences in preference weights as well as RIS attribute rankings between DCE and BWS-2. Latent class analysis revealed similar latent classes for the same sample of patients. Patients also did not find DCE more difficult compared to BWS-2.

In the general discussion (**chapter 10**), the main findings of **chapters 2-9** are integrated and further discussed. This dissertation explores and addresses several methodological challenges and opportunities for choice modelling – specifically DCE and BWS-2 – in health. While these methods have the potential to provide useful choice evidence to inform health decision-making, they also provide decision-makers with an additional source of information that might actually complicate the decision process for policy makers. Ultimately, this dissertation adds to the literature aiming to provide new insights how choice modelling can provide useful information for decision-making, although more research is needed regarding the generalizability (external validity) of choice modelling outcomes and its relevance for which type of decision-making process.

Samenvatting

Het veld van de keuzemodellering houdt zich bezig met het begrijpen hoe individuen keuzes maken door de onderliggende voorkeuren te kwantificeren. Meer specifiek is het doel van keuzemodellering om individuele keuzes te karakteriseren en om keuzes te voorspellen op basis van de relevante keuze alternatieven. Keuzemodelleurs gaan ervan uit dat keuzes van individuen gebaseerd zijn op voorkeuren die de mate van voldoening (utiliteit) bepalen die ze ontlenen aan goederen en diensten. In het veld van keuzemodellering worden de keuzes van individuen gerelateerd aan hun voorkeuren door zich te richten op de utiliteiten van de keuze alternatieven. Keuzemodellering (in dit geval specifiek discrete keuze-experimenten) werd begin jaren negentig in de gezondheidszorg geïntroduceerd, met name om uitkomsten te meten die verder reiken dan alleen gezondheid in de context van *health benefit assessments*. Voorafgaand aan de introductie van keuzemodellering in de gezondheidszorg werden methoden uit andere vakgebieden gebruikt om inzicht te krijgen in gezondheidsvoorkeuren. Omdat keuzemodellering een nieuwe manier bood om inzicht te krijgen in deze voorkeuren, bleef de toepassing van keuzemodellering in de gezondheidszorg na de jaren negentig groeien, vooral met betrekking tot de zogeheten *stated-preferences*.

Er zijn verschillende *stated-preference* methoden om inzicht te krijgen in gezondheidsvoorkeuren, maar er wordt steeds meer gepleit voor het gebruik van discrete keuze-experimenten (DCE). Een DCE is een op een vragenlijst gebaseerde methode voor het meten van voorkeuren waarbij individuen wordt gevraagd hun keuze alternatief te selecteren uit een reeks van voorgelegde hypothetische alternatieven. DCE data-analyse vindt zijn oorsprong in de mathematische psychologie, met brede toepassingen in marketing, transport en milieu-economie en zijn theoretische basis in de *random utility theory* (RUT). *Best-worst scaling* (BWS) is een andere steeds populairder wordende methode om gezondheidsvoorkeuren te meten. De introductie van BWS kwam voort uit de behoefte om meer informatie over voorkeuren te verkrijgen in vergelijking met een DCE, door individuen te vragen hun "beste" en "slechtste" optie te selecteren zonder daarbij de cognitieve belasting te vergroten. Er zijn verschillende soorten BWS, maar *case 2 BWS* (BWS-2) heeft veel aandacht gekregen in de academische literatuur. Deze methode kan namelijk inzicht geven in het belang van het attribuutniveau, de cognitieve belasting van keuzetaken verminderen door zich op één profiel tegelijk te concentreren en deze keuze-experimenten zijn relatief makkelijk te ontwerpen. Meer gedetailleerde informatie over deze methoden is te vinden in **hoofdstuk 1** van dit proefschrift.

DCE en BWS-2 zijn populaire methoden om gezondheidsvoorkeuren te meten, hoewel er verschillende methodologische uitdagingen zijn die verhinderen dat deze methoden nog waardevoller kunnen zijn voor de daadwerkelijke besluitvorming. Om deze reden draagt het verkrijgen van inzichten in deze methodologische uitdagingen en kansen om deze op te lossen bij aan zowel de academische literatuur als de praktijk van de besluitvorming. Dit proefschrift kent drie hoofddoelen (**hoofdstuk 1**):

1. Inzicht geven in methoden die in de gezondheidszorg gebruikt worden om voorkeuren te meten
2. Inzicht geven in de uitdagingen en kansen van DCE en BWS-2 met betrekking tot ontwerp en analyse
3. Empirische vergelijking van DCE en BWS-2 uitkomsten

Inzicht in de methoden in de gezondheidszorg om voorkeuren te meten

Hoofdstuk 2 biedt een actueel compendium en taxonomie van zowel *preference exploration* (kwalitatieve) als *preference elicitation* (kwantitatieve) methoden, aangezien het belangrijk is om onderzoek naar de integratie van voorkeuren in gezondheidsbeslissingen verder te stimuleren. Een drie-stappenplan is gebruikt om bestaande methoden om voorkeuren in kaart te brengen te identificeren. Ten eerste is een systematische literatuurstudie uitgevoerd waarbij 4572 unieke artikelen zijn geïdentificeerd via verschillende wetenschappelijke databases, gebruik makend van Engelse *full-text* artikelen die tussen 1980 en 2016 zijn gepubliceerd. Ten tweede zijn eerdere literatuurstudies over methoden om voorkeuren in de gezondheidszorg in kaart te brengen onderzocht om deze resultaten te valideren. Ten derde werden internationale experts (n=24) geraadpleegd om deze resultaten te bevestigen en om andere mogelijke methoden te identificeren. De systematische literatuurstudie (19 methoden), analyse van eerder uitgevoerde studies (23 methoden) en expertconsultaties (4 methoden) hebben allemaal bijgedragen aan het compendium. In totaal werden 32 unieke methoden geïdentificeerd. Deze methoden zijn gegroepeerd in verschillende categorieën om een actueel compendium en taxonomie te bieden.

Hoofdstuk 3 geeft een overzicht en beschrijving van trends, de huidige praktijk en toekomstige uitdagingen van DCE toepassingen en methoden in de gezondheidseconomie sinds drie eerder gepubliceerde systematische literatuuronderzoeken (welke periode 1990-2012 bestrijken). Deze literatuurstudie geeft informatie of de uitdagingen die in eerdere studies zijn geïdentificeerd nog steeds relevant zijn of dat er een reactie is geweest op de gepubliceerde suggesties en richtlijnen, aangezien de belangrijkste belemmeringen voor een breder gebruik van DCE's in beleid betrekking hadden op de robuustheid en validiteit van de methode en de kwaliteit van toegepaste studies. Tussen 2013-2017 werden 301 studies geïdentificeerd die voldeden aan de inclusiecriteria en waarvan de gegevens werden geëxtraheerd. De resultaten toonden aan dat het totale aantal DCE's per jaar bleef toenemen, met bredere toepassingsgebieden en een grotere geografische reikwijdte. Studies rapporteerden ook het gebruik van meer geavanceerde DCE ontwerpen en het gebruik van specifieke software om deze te genereren. De trend van het gebruik van meer geavanceerde econometrische modellen zette zich ook door. Veel studies presenteerden echter geavanceerde methoden met onvoldoende detail. Bovendien bleef kwalitatief onderzoek een populaire manier om attribuut en attribuutniveaus te identificeren. Een gebrekkige rapportage van methodologische details kan echter een belemmering vormen voor de kwaliteitsbeoordeling en kan daarbij het vertrouwen van de besluitnemer om te handelen naar DCE uitkomsten verminderen.

Inzicht in de uitdagingen en kansen van DCE en BWS-2 met betrekking tot ontwerp en analyse

Hoofdstukken 4-7 van dit proefschrift omvatten DCE en BWS-2-studies gericht op methodologische uitdagingen en kansen met betrekking tot het ontwerp van DCE en BWS-2 experimenten, evenals de analyse van DCE en BWS-2 keuzedata. In **hoofdstuk 4** werd de nauwkeurigheid van voorspellingen van DCE *choice shares* onderzocht: onderzoek of deze voorspellingen af zouden hangen van de gebruikte econometrische modelleringsaanpak en het type analyse dat zou worden uitgevoerd, rekening houdend met de heterogeniteit in voorkeuren. De resultaten van de simulatiestudie gaven aan dat modellen die wel en geen rekening hielden met heterogeniteit in voorkeuren, kunnen worden gebruikt om schattingen te verkrijgen en *choice shares* te voorspellen. Echter, een model dat helemaal geen rekening houdt met heterogeniteit in voorkeuren (*multinomial logit*, MNL) presteert beter (d.w.z. nauwkeurigere voorspellingen) in vergelijking met een meer gecompliceerd model dat wel rekening houdt met heterogeniteit (*mixed logit*, MXL) wanneer heterogeniteit wordt genegeerd in de voorspellingsfase. De nauwkeurigheid van de voorspellingen van DCE *choice shares* wordt gemaximaliseerd bij gebruik van MXL en daarbij rekening houdend met heterogeniteit in voorkeuren in de voorspellingsfase. Deze uitkomsten hebben laten zien dat het uitvoeren van niet-corresponderende analyses in de voorspellingsfase een enorme impact heeft op de uitkomsten en daarmee op besluitvorming.

In **hoofdstuk 5** wordt de prestatie van BWS-2 bestudeerd wanneer het wordt toegepast bij een mix van positieve en negatieve attributen (bijv. het bestuderen van behandelingen die worden gekenmerkt door zowel voordelen als nadelen). Deze mix leidt intuïtief tot dominantie van attributen. Analytisch is aangetoond dat deze dominantie leidt tot oneindig grote verschillen tussen de parameterschattingen voor de positieve versus de negatieve attributen en daarmee tot schattingsproblemen. De simulatiestudie bevestigde onze analytische resultaten: parameterwaarden van de attributen konden niet nauwkeurig worden teruggevonden. Wanneer slechts één enkel positief attribuut werd gebruikt, werd zelfs de relatieve volgorde van de attribuutniveau voorkeuren niet geïdentificeerd. **Hoofdstuk 6** bood een mogelijke oplossing voor het probleem van het mixen van positieve en negatieve attributen in BWS-2. In deze studie werden attributen ofwel allemaal positief, allemaal negatief of gemengd gedefinieerd. In totaal werden 192 patiënten geïncludeerd voor analyse, waarbij er verschillen in uitkomsten tussen positief en negatief gedefinieerde BWS-2 werden gevonden. De resultaten toonden ook aan dat het negatief definiëren er toe leidde dat attributen belangrijker werden, terwijl het positief definiëren van attributen er juist toe leidde dat attributen minder belangrijk werden. Deze studie leverde ook bewijs dat het mengen van positieve en negatieve attributen in BWS-2 leidt tot theoretisch vreemde resultaten.

Het laatste hoofdstuk van deze sectie (**hoofdstuk 7**) presenteert een studie die tot doel had een nieuwe BWS-2 benadering te introduceren, met expliciete referentiepunten in de keuzetaken (BWS-2R) en te onderzoeken of deze nieuwe benadering leidde tot een meer accurate analyse van voorkeuren vergeleken met BWS-2. Analytisch werd aangetoond dat BWS-2R de ruis in de voorkeuren zou moeten verminderen, aangezien het meten van voorkeuren via reguliere BWS-2 hinder ondervindt van referentiepunten. Onze empirische onderzoeksresultaten (n=601) toonden statistisch significante verschillen tussen geschatte voorkeuren voor beide BWS-2 benaderingen, evenals statistisch significante verschillen in *relative importance scores* (RIS) tussen BWS-2R en BWS-2. Er werd geen verschil gevonden in moeilijkheid tussen de twee benaderingen, waarbij een groter deel van de respondenten die BWS-2 voltooiden de voorkeur gaf aan BWS-2R dan andersom.

Empirische vergelijking van DCE en BWS-2 uitkomsten

Hoofdstuk 8 presenteert een empirische studie waarin de ervaren cognitieve belasting tussen DCE en BWS-2 wordt vergeleken. Resultaten van een gerandomiseerd experiment (n=469) toonden aan dat bij het gebruik van verschillende visuele presentaties van DCE en BWS-2-keuzetaken (d.w.z. attribuutniveau-overlap en kleur codering), DCE-taken minder cognitief belastend bleken te zijn in vergelijking met BWS-2. Vooral kleur codering in BWS-2 kon niet worden aanbevolen, omdat het effect op de cognitieve belasting niet duidelijk was en leidde tot ongewenste keuzeheuristieken. In **hoofdstuk 9** werden DCE en BWS-2 empirisch vergeleken in termen van statistische uitkomsten. Resultaten van deze empirische studie (n=140) wezen op verschillen in voorkeuren en RIS rangschikkingen tussen DCE en BWS-2. Latente klassenanalyse onthulde vergelijkbare latente klassen voor dezelfde steekproef van patiënten. Patiënten vonden DCE ook niet moeilijker in vergelijking met BWS-2.

In de algemene discussie (**hoofdstuk 10**) worden de belangrijkste bevindingen van de **hoofdstukken 2-9** geïntegreerd en verder besproken. Dit proefschrift onderzoekt en behandelt verschillende methodologische uitdagingen en kansen voor keuzemodellering – met name DCE en BWS-2 – in de gezondheidszorg. Hoewel deze methoden het potentieel hebben om bruikbaar bewijs te leveren voor besluitvorming in de gezondheidszorg, bieden ze besluitnemers ook een aanvullende informatiebron die het besluitvormingsproces voor bijvoorbeeld beleidsmakers zou kunnen compliceren. Uiteindelijk draagt dit proefschrift bij aan de literatuur met als doel nieuwe inzichten te verschaffen hoe keuzemodellering nuttige informatie kan opleveren voor besluitvorming, alhoewel er meer onderzoek nodig is naar de generaliseerbaarheid (externe validiteit) van keuzemodellering uitkomsten en de relevantie ervan voor welk type besluitvormingsproces.

All publications

- Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto C, Donkers B, Huys I, van Overbeeke E, Juhaeri J and de Bekker-Grob EW. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. Drug Discovery Today. 2019;24(7):1324-1331. doi:10.1016/j.drudis.2019.05.001

- Soekhai V, de Bekker-Grob EW, Ellis AR and Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019;37(2):201-226. doi:10.1007/s40273-018-0734-2

- Donkers B, Soekhai V, Levitan B and de Bekker-Grob EW. Predicting choice shares with discrete choice experiments: using the right model and conducting the right analysis. [submitted]

- Soekhai V, Donkers B, Levitan B, de Bekker-Grob EW. Case 2 best-worst scaling: For good or for bad but not for both. Journal of Choice Modelling. Published online 2021:100325. doi:https://doi.org/10.1016/j.jocm.2021.100325

- Soekhai V, Donkers B, van Kinschot CMJ, van Noord C and de Bekker-Grob EW. Framing Attribute Levels in Case 2 Best-Worst Scaling: Do signs matter? [submitted]

- Soekhai V, Donkers B, de Bekker-Grob EW. Implicit versus Explicit Reference Points in Case 2 Best-Worst Scaling Tasks [submitted]

- Himmler S, Soekhai V, van Exel J and Brouwer W. What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. Journal of Choice Modelling. 2021;38:100265. doi:10.1016/j.jocm.2020.100265

- Soekhai V, Donkers B, Viberg Johansson J, Jimenez-Moreno C, Pinto CA, de Wit GA, de Bekker-Grob EW. Comparing DCE and BWS-2 Outcomes: an application to Neuromuscular Disease treatment. [submitted]

- Cecilia Jimenez-Moreno A, Anne Pinto C, Levitan B, Whichello C, Dyer C, van Overbeeke E, de Bekker-Grob EW, Smith I, Huys I, Viberg Johansson J, Adcock K, Bullock K, Soekhai V, Yuan Z, Lochmuller H, de Wit A and Gorman GS. A study protocol for quantifying patient preferences in neuromuscular disorders: a case study of the IMI PREFER Project Published online 2020. doi:10.12688/wellcomeopenres.16116.1

- Collacott H, Soekhai V, Thomas C, Brooks A, Brookes E, Lo R, Mulnick S and Heidenreich S. A Systematic Review of Discrete Choice Experiments in Oncology Treatments. Patient - Patient-Centered Outcomes Research 2021. Published online May 5, 2021:1-16. doi:10.1007/S40271-021-00520-4

- Kinschot CMJ van, Soekhai V, Bekker-Grob EW de, Visser EW, Peeters RP, van Ginhoven TM and van Noord C. Preferences of patients and clinicians for treatment of Graves' disease: a discrete choice experiment. European Journal of Endocrinology. 2021;184(6):803-812. doi:10.1530/EJE-20-1490

PhD portfolio

PhD candidate	Vikas Rogier Soekhai
Affiliation	Erasmus School of Health Policy & Management, Erasmus University Rotterdam Department of Public Health, Erasmus MC, University Medical Center
Promotors	Prof.dr. E.W. de Bekker-Grob Prof.dr. B. Donkers
PhD Period	2016 -2021

	Year	ECTS
Training activities		
Courses		
Julia Scientific Programming	2016	2.0
Scientific Writing in English for Publication	2016	1.0
Econometrie 1	2017	4.0
Econometrie 2	2017	4.0
Bayesian Econometrics (audited)	2017	4.0
Academic integrity training	2017	0.2
PhD career day	2017	0.2
Measurement of Patient Preferences using Discrete Choice Experiments (audited)	2017	5.0
Choice modelling and survey design	2018	1.4
R programming	2018	2.0
Health Service Operations Management (audited)	2019	5.0
Python Bootcamp	2019	0.2
PhD jobmarket	2020	0.2
Seminars		
Seminars department of Public Health Erasmus MC	2016-2018	0.5
Meetings Medical Decision Making Erasmus MC	2016-2018	0.3
Seminars Erasmus Choice Modelling Centre EUR	2018-2020	0.5
Seminars Health Economics EUR	2018-2020	0.1
Seminars Health Technology Assessment EUR	2018-2020	0.1
National and international conferences		
Random Regret Minimization Symposium	2016	0.2
5th International Choice Modelling Conference	2017	1.0
ISPOR 20th Annual European Congress	2017	1.0
7th Meeting of The International Academy of Health Preference Research	2017	0.4
2017 Lowlands Health Economics Study Group Conference	2017	0.6
2018 Lowlands Health Economics Study Group Conference	2018	0.6
18th Biennial European Conference Society of Medical Decision Making	2018	0.5
ISPOR 21st Annual European Congress	2018	1.0
10th Meeting of The International Academy of Health Preference Research	2019	0.4
2019 Lowlands Health Economics Study Group Conference	2019	0.6
ISPOR 22nd Annual European Congress	2019	1.0
The Cancer Drug Development Forum Multi-takeholder Meeting	2019	0.2
Teaching activities		
Assistance master thesis supervision	2018-2019	0.1
Master thesis supervision (3 students)	2019-2020	4.5
Interprofessional eduction ESHPM-EMC: How to keep healthcare affordable?	2019	0.2
Workgroup teacher Measurement of Patient Preferences using Discrete Choice Experiments	2018	0.3
Workgroup teacher Measurement of Patient Preferences using Discrete Choice Experiments	2019	0.3
Workgroup teacher Statistics	2019	0.4
Visiting lecturer Pontificia Universidad Javeriana Bogotá Colombia: DCEs, Theory and Practice	2020	1.5
Other activities		
Peer reviewer scientific publication (1)	2018	0.1
Peer reviewer scientific publication (3)	2019	0.3
Peer reviewer scientific publication (3)	2020	0.3
PREFER Consortium Annual meetings	2016-2019	2.5
TOTAL		**48.7**

Dankwoord

Een aantal jaar geleden stond ik na een periode in de consultancy gewerkt te hebben voor de mogelijkheid om te starten aan dit promotieonderzoek. Het was een lastige keuze, aangezien het verrichten van wetenschappelijk onderzoek zeker mijn aandacht trok, maar ik door de verschillende ervaringen in mijn omgeving met promotieonderzoek niet goed wist wat ik zou kunnen verwachten. Achteraf ben ik toch blij dat ik heb gekozen voor dit promotieonderzoek: het waren afwisselende, uitdagende, leerzame en kansrijke jaren. Ik wil daarom ook iedereen bedanken die op zijn of haar manier heeft bijgedragen aan de totstandkoming van dit proefschrift.

Mijn promotoren: Esther de Bekker-Grob en Bas Donkers. Esther, ik waardeer enorm het vertrouwen dat je in mij hebt gehad, evenals je inhoudelijk sterke alsmede pragmatische aanpak. Ik heb mij ook altijd afgevraagd hoe je zoveel ballen tegelijkertijd in de lucht kunt houden en nog steeds altijd de juiste vragen kon stellen of oplossingen kon bedenken: enorm veel respect voor! Ik wil oprecht zeggen dat ik trots ben dat ik onder jouw vleugels dit proefschrift heb kunnen schrijven. Dank voor de fijne samenwerking, iets wat ik zeker ga missen, heel veel succes in je nieuwe rol als hoogleraar en het gaat je goed. Bas, ook jij hartelijk dank voor de fijne samenwerking in de afgelopen jaren. Ik heb je kritische blik in alle eerlijkheid zo af en toe vervloekt, maar uiteindelijk heb ik altijd eerlijk moeten toegeven dat het werk er uiteindelijk altijd beter door is geworden. Dank dat ik altijd een beroep op je kon doen voor de technisch ingewikkelde vragen waar je altijd een mooi antwoord op had en heb je toegankelijkheid ook altijd erg gewaardeerd. Dankjewel!

Ik wil daarnaast ook nog Ewout Steyerberg bedanken. Ewout, aanvankelijk was jij één van mijn promotoren uit het Erasmus MC. Echter, door je vertrek uit Rotterdam naar Leiden en mijn overstap van het Erasmus MC naar de EUR, zijn onze lijntjes niet meer zo kort gebleven als gehoopt. We hebben in het begin van het proefschrifttraject leuke gesprekken gehad over verschillende zaken, maar de bitcoin discussies blijven mij toch bij (hopelijk ben je nog steeds aan het HODL-en). Ik ben blij dat je toch nog onderdeel wilde zijn van mijn promotiecommissie. Ik wil hierbij direct de leden van zowel de kleine als grote promotiecommissie bedanken voor de tijd die zij hebben besteed aan mijn proefschrift en voor het opponeren tijdens de proefschrift verdediging.

Mijn dank gaat ook uit naar de collega's van het Erasmus Choice Modelling Centre. Een mooie groep van kundige en prettige collega's waar laagdrempelig altijd (nieuwe) ideeën getoetst konden worden. Dank Jorien voor je bereidheid om altijd te helpen bij (DCE gerelateerde) vragen en dank Marcel voor het altijd kritische meedenken. Also thank you to all IMI-PREFER colleagues: I really enjoyed working together with you on such a large international project. Especially thanks to Bennett Levitan and Juhaeri Juhaeri for our collaboration on the simulation studies. Tenslotte wil ik alle co-auteurs bedanken voor hun waardevolle inbreng en de prettige samenwerking.

In het bijzonder wil ik daarnaast nog mijn kamergenootjes van de gezelligste kamer op de 8^{e} verdieping van het ESHPM gebouw bedanken. Samare, je enthousiasme, humor en relativeringsvermogen zijn erg mooi om te zien. En ik was erg blij dat ik nu tenminste in mijn directe omgeving iemand kon lastigvallen met allerlei choice modelling vragen. Laurenske, je overstap naar onze kamer paste precies en zorgde er alleen maar voor dat het PhD-leven nog meer kleur kreeg. Frederick, onze R connaisseur. Erg leuk om met jou een kamer gedeeld te hebben en ben trots op je dat je je proefschrift succesvol hebt kunnen afronden eind vorig jaar. Het was voor mij zoeken naar de juiste kamer na mijn overstap, maar ben erg blij dat ik bij jullie op de kamer terecht ben gekomen (al was het alleen maar voor de cola breaks en de wandeling richting de SPAR). Al heeft COVID ervoor gezorgd dat er een wat abrupt einde kwam aan de gezelligheid op de kamer. Succes met alles en we komen elkaar vast weer eens tegen.

Joost en Wesley, wij kennen elkaar al sinds de middelbare school. Ook al zijn er periodes dat we elkaar een tijdje niet zien, is het toch altijd weer van ouds wanneer we elkaar weer spreken. Dank voor de nodige gezelligheid! Ik wil ook mijn oud-DSW collega's bedanken. Jasper en Jeroen: zonder de voetbalavondjes (inclusief het bijzondere Champions League seizoen van een club die in Rotterdam niet zo goed ligt), FIFA avondjes en etentjes zou het een stuk lastiger zijn geweest dit proefschrift tot een einde te brengen. Daarnaast wil ik Marko en Ilija, mijn oud-studiegenootjes van de rechtenopleiding, bedanken voor de gezellige etentjes en borrels in de afgelopen jaren. Al is de levenssituatie van ons allen veranderd, ik hoop dat we elkaar zeker nog zullen blijven zien. Tenslotte Oemar, ik kwam je tegen in de gang bij ESHPM en dacht ik ga jou gewoon eens vragen een keer een kop koffie te doen. Nu een aantal jaar verder hebben we samengewerkt, in verschillende vormen, aan projecten over onze gezamenlijke passie: de Surinaamse gezondheidszorg. Met als voorlopig hoogtepunt een uitnodiging en bezoek aan de Surinaamse ambassadeur in Nederland. We gaan elkaar zeker nog zien en spreken. Heel veel succes met het afronden van je proefschrift.

Lieve papa en Linda: bedankt voor jullie steun en interesse de afgelopen jaren. Ik weet dat het voor jou papa door het verlies van mama niet makkelijk is geweest. Ik doe het je niet na: in een vreemd land een jong kind opvoeden. Maar ik kan voor de volle honderd procent zeggen dat mama trots op je zou zijn geweest, je had het niet beter kunnen doen. Dank voor je geloof in mij en ik hoop dat ik je een beetje trots heb kunnen maken. Linda, jij hebt je altijd als een moeder over mij ontfermt en zo heb ik dat ook altijd gevoeld. Ondanks je niet te beste gezondheid heb je altijd alles voor mij gedaan en ook altijd geloof in mij gehad. Dank voor alles, ook al laat ik het niet altijd even goed blijken. Ik wil daarnaast ook de rest van de (schoon)familie in zowel Nederland als Suriname hartelijk danken. Mijn “studie” is nu eindelijk afgerond en heb nu een “echte” baan!

Nick, wij kennen elkaar al sinds de basisschool (en volgens onze ouders al daarvoor bij de zwemlessen, maar dat kan ik mij niet al te best herinneren). Ik heb je altijd beschouwd als mijn beste vriend en ik ben erg blij dat we nog zulk goed contact hebben gehouden en altijd over van alles en nog wat kunnen praten. Je gaat een mooi hoofdstuk in je leven openen met een kleintje op komst, heel veel plezier, geluk en dank dat je mijn paranimf wilde zijn.

Lieve Kawita, de final words zijn voor jou. Dank voor je support in de afgelopen jaren en het aanhoren van al mijn proefschrift verhalen. Daar wist je als neuropsycholoog altijd wel goed raad mee. We hebben fijne jaren gehad tijdens dit promotietraject, vooral de vakanties die we altijd achter een buitenlands congres probeerden te plakken. Soms is het ook lastig geweest, maar alles is altijd weer op z’n plekje gevallen. We zijn een goed team samen, al zeg ik het zelf, en ik kijk uit naar de volgende fases in ons leven samen. Dank dat je mijn paranimf wilde zijn.

Vikas

About the author

Vikas Rogier Soekhai was born on the 1st of May 1989 in Leiden, the Netherlands. In 2007 he finished secondary school (Segbroek College, Den Haag). In 2011 Vikas obtained his master's degree in Health Economics (Erasmus University Rotterdam) and in 2014 he received his master's degree in Health Law (Erasmus University Rotterdam).

After working several years inside (researcher) and outside (consultant) healthcare, Vikas started his PhD trajectory on choice modelling in health late 2016 at the Erasmus University Rotterdam (Erasmus School of Health Policy & Management) and Erasmus MC, University Medical Center (Department of Public Health). His research was part of the Erasmus Choice Modelling Centre, which involves researchers from the Erasmus School of Economics, Erasmus MC, University Medical Center and Erasmus School of Health Policy & Management. Furthermore, his research was also part of the PREFER project: an public-private collaborative research project under the Innovative Medicines Initiative (IMI) which aims to strengthen the role of patient preferences in the medical product lifecycle.

As of January 2021, Vikas is working as a policymaker at the Dutch Ministry of Health, Welfare and Sport.