


Using Best–Worst Scaling to Investigate Preferences in Health Care

Kei Long Cheung¹  · Ben F. M. Wijnen^{1,2} · Ilene L. Hollin³ · Ellen M. Janssen³ · John F. Bridges³ · Silvia M. A. A. Evers¹ · Mickael Hiligsmann¹

Published online: 11 July 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Introduction Best–worst scaling (BWS) is becoming increasingly popular to elicit preferences in health care. However, little is known about current practice and trends in the use of BWS in health care. This study aimed to identify, review and critically appraise BWS in health care, and to identify trends over time in key aspects of BWS.

Methods A systematic review was conducted, using Medline (via Pubmed) and EMBASE to identify all English-language BWS studies published up until April 2016. Using a predefined extraction form, two reviewers independently selected articles and critically appraised the study quality, using the Purpose, Respondents, Explanation, Findings, Significance (PREFS) checklist. Trends over time periods (≤ 2010 , 2011, 2012, 2013, 2014 and 2015) were assessed further.

Results A total of 62 BWS studies were identified, of which 26 were BWS object case studies, 29 were BWS profile case studies and seven were BWS multi-profile case studies. About two thirds of the studies were performed in the last 2 years. Decreasing sample sizes and decreasing numbers of factors in BWS object case studies, as well as use of less complicated analytical methods, were observed in recent studies. The quality of the BWS studies was generally acceptable according to the PREFS checklist, except that most studies did not indicate whether the responders were similar to the non-responders.

Conclusion Use of BWS object case and BWS profile case has drastically increased in health care, especially in the last 2 years. In contrast with previous discrete-choice experiment reviews, there is increasing use of less sophisticated analytical methods.

K. L. Cheung and B. F. M. Wijnen contributed equally to this work and share co-first authorship.

Electronic supplementary material The online version of this article (doi:10.1007/s40273-016-0429-5) contains supplementary material, which is available to authorized users.

✉ Kei Long Cheung
KL.Cheung@maastrichtuniversity.nl

¹ Department of Health Services Research, CAPHRI Research School for Public Health and Primary Care, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

² Department of Research and Development, Epilepsy Centre Kempenhaeghe, Heeze, The Netherlands

³ Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Key Points for Decision Makers

This systematic review identified 62 best–worst scaling (BWS) studies conducted in health care and published up until April 2016. About two thirds of the studies were performed in the last 2 years, indicating the increasing popularity of the method.

BWS is an attractive and relatively easy method to investigate preferences over a wide range of health care topics.

Most BWS studies in this review were of acceptable quality according to the PREFS checklist. However, researchers should give more attention to reporting whether responders are similar to non-responders in a study. This may reduce the risk of bias and increase the generalisability of the findings.

1 Background

Investigating preferences has become popular in health care. In recent years, the need for a more patient-centred approach in health care and health technology assessment has been widely acknowledged. Hence, it is expected that eliciting preferences will become increasingly important to support health policy and clinical decisions [1, 2]. In order to enhance the patient-centred approach, preferences may be elicited from patients or other key players in the health sector, such as clinicians and policy makers, as to how they value various aspects of health interventions or when designing and evaluating health care programmes [3]. By eliciting the preferences of other key players in the health sector, one may identify potential discrepancies between the stakeholders, which may be addressed to enhance patient-centred health care [4]. Preference studies can provide relevant information to support health technology assessment and decisions about priorities in health care [5]. Better understanding of patients' preferences for treatment and involvement of patients in clinical decision making could also help to optimise disease management and medication adherence.

Several methods have been developed and used to assess preferences in health care, ranging from simple rating scale exercises to more complex forms of preference valuation techniques, such as conjoint analysis. Conjoint analysis is a decomposition method, which derives implicit values for factors (or attributes and levels) from some overall score for a profile consisting of two or more factors [6, 7]. In particular, the use of traditional discrete-choice experiments (DCEs) has increased drastically in recent years [2, 8]. Although DCEs may take many forms, this study refers solely to the 'traditional' DCE in which a respondent typically chooses one alternative among a number of competing multi-profile options described by varied factors. Another type of conjoint analysis—namely, best-worst scaling (BWS)—is becoming increasingly popular in health care [9]. BWS—which was devised by Finn and Louviere [10], first applied in health care by Szeinbach et al. [11] (who named it 'maximum difference conjoint analysis' [11]) and formally introduced to health care research by McIntosh and Louviere [12]—differs from traditional DCEs, mainly because it elicits additional information on the least preferred option [13]. BWS consists of choice tasks, with a minimum of three items/options, in which a person is asked to indicate the best and the worst items/options, with the overall aim being to obtain a full ranking of items in a manner that is easy for respondents and can then be analysed in various ways [9].

As health care applications of BWS have emerged only recently, it is important to reflect on and assess current

practice. Several reviews of DCEs have been conducted [14, 15] and revealed trends in the conduct and use of DCEs. For BWS studies, although they have not been systematically reported, scholars have noted the increasing popularity of BWS in health care [13, 16]. Yet, little is known about current practice and trends in the use of BWS in health care, except from two recent overview articles of BWS studies by Mühlbacher et al. [17, 18]. These articles provided insights into the possibilities of application, the theoretical concepts underlying the BWS method, analytical methods, strengths and limitations of BWS case types and lists of studies that have applied the BWS method. However, these articles did not report on current practice by BWS case type, quality appraisal of BWS studies and trends in BWS studies in health care. Reviewing the use of BWS and reflecting on current practice could help to improve best practice of BWS and potentially lead to increased interest from policy makers or physicians in the wide usability and value of BWS. Hence, we conducted a systematic review to identify and analyse all BWS studies conducted in health care, and to identify if key aspects of BWS have changed over time.

2 Best–Worst Scaling

BWS can be used to determine preferences for a wide range of health care questions, by asking the respondent to indicate the best and the worst in a set of available items or options. There are three types of BWS, which differ in terms of the complexity of the items or options under consideration: BWS object case (case 1), BWS profile case (case 2) and BWS multi-profile case (case 3) [16].

In BWS object case, the researcher is interested in the relative values associated with each of a list of objects. Figure 1a illustrates the case of three relevant attributes. Once the researcher has chosen the list of objects, he or she presents choice sets of these to respondents to obtain best and worst option data. Here, the objects are not decomposed into factors. BWS object case can be used, for instance, to examine features of a new medicine. Hence, one needs to develop a list of mutually exclusive objects [19]. BWS object case can be very attractive in health care because health care goods/services can be complicated, and even pairs of specifications (e.g. a simple DCE) may lead to an unacceptable cognitive burden, particularly among vulnerable patient groups [16].

In BWS profile case—in contrast to BWS object case—the level of each attribute is shown, and the choice set has the structure of a single profile. In Fig. 1b, a possible treatment is characterised by five attributes, providing its levels in each choice set. However, the respondent is not

Fig. 1 Examples of **a** best–worst scaling (BWS) object case, **b** BWS profile case and **c** BWS multi-profile case

a	Least important	Factors of pharmaceutical	Most important
	X	Side-effects	
		Frequency of intake	X
		Mode of administration	

b	Least important	Factors of pharmaceutical	Most important
		Side effects: moderate	
		Frequency of intake: daily	
		Mode of administration: oral supplement	X
	X	Out-of-pocket expenses: €10 per week	
		Duration of treatment: eight weeks	

c	Factors	Drug A	Drug B	Drug C
	Side effects:	Moderate	Severe	Severe
	Frequency of intake:	Daily	Monthly	Monthly
	Mode of administration:	Oral supplement	Injection	Injection
	Out-of-pocket expenses:	€10 per week	€15 per week	€5 per week
	Duration of treatment:	Eight weeks	Twelve weeks	Four weeks
	Which drug would be the best treatment for you?	X		
	Which drug would be the worst treatment for you?			X

required to consider the value of the profile as a whole, but must consider the factor levels that describe it, and must choose the one that is best (most attractive) and the one that is worst (least attractive). Hence, profiles must describe particular configurations or specifications of goods or a service, in which each profile represents a combination of factors and associated levels [19]. The factors are fixed across all profiles, but the combinations of factor levels differ by profile.

In contrast to BWS profile case, BWS multi-profile case provides multiple profiles, and the respondent is asked to choose the best (most attractive) profile, as well as the worst (least attractive) one, in the current choice set. Figure 1c illustrates a choice set, again using a possible treatment as an example. Respondents base their choice on the whole scenario instead of single factor levels, as in BWS profile case. The multiple profiles include all of the factors and one level per factor, and the respondent is assumed to make deliberate choices based on the presented levels for each factor. BWS multi-profile case is most similar to a traditional DCE but extends the general DCE design to allow for best and worst choices [19].

3 Methods

3.1 Literature Search

A systematic review was conducted, using Medline (via Pubmed) and EMBASE to identify all English-language BWS studies published in print or online between January 1990 and April 2016. In addition, bibliographies of BWS applications identified in the literature search were searched manually (forwards and backwards) for additional articles. Following the examples of DCE reviews from Ryan and Gerard [2] and Clark et al. [14], the search strategy was formulated according to free-text terms. The following free-text terms focusing on BWS studies were used individually in each database: ‘object scaling’ OR ‘BWS’ OR ‘best worst scaling’ OR ‘best worst’ OR ‘max diff’ OR ‘maxdiff’ OR ‘max diff scaling’, ‘maxdiff scaling’ OR ‘maximum difference scaling’ (see Electronic Supplementary Material 1). The research team discussed and approved the list of free-text terms. In addition, we cross-checked the identified articles with the articles identified in the recent review by Mühlbacher et al. [17].

3.2 Selection of Studies

Titles and abstracts were screened independently by two researchers (KLC and BFMW). Studies were included if they were original BWS studies, were published as full-text articles and applied to health care. The definition published by Louviere et al. [19] was used for inclusion of BWS studies, meaning that only BWS studies in which respondents needed to indicate both a best and a worst option per choice task were included. Hence, studies looking at best or worst alternatives only, and studies using only best data to estimate models, were excluded. Furthermore, studies based on simulated data were excluded. Disagreements about study inclusion or exclusion based on the abstract review were resolved by consensus. At least two independent reviewers (among KLC, BFMW, ILH, EMJ) then reviewed the full texts of all remaining articles. Consensus on inclusion of each study was reached by at least two reviewers.

3.3 Data Extraction

For each article that met our inclusion criteria, two independent reviewers (KLC and EMJ, and BFMW and ILH) systematically extracted data by using a predefined extraction form, which included a variety of data elements relevant to the conduct of BWS experiments. Any disagreements were resolved by consensus. This data extraction form was based on a review by de Bekker-Grob et al. [20] on DCE and was further developed through discussion within the research team. A pilot study was performed by four reviewers for four studies, which led to some adjustments to the data extraction form. The extracted data included three categories: general study information, design and choice sets, and estimation procedure and validity. General study information included the authors, year of publication, study name, study type, study setting, target population, type of case and area of application. Design and choice sets included the sample size, methods for factor selection, type of BWS, design of BWS, domains of BWS factors, number of best–worst options per choice set, number of included factors, maximum number of choice tasks, number of factors per choice task, BWS survey administration and BWS survey method. Estimation procedure and validity included the analytical method, use of software packages to analyse BWS experiments, preference heterogeneity taken into account (by means of latent class or random parameters analyses), subgroup analysis and use of pilot validity checks (i.e. test–retest, face validity of a dominant choice or subgroup analysis).

3.4 Quality Assessment

In order to assess the study quality, the Purpose, Respondents, Explanation, Findings, Significance (PREFS) checklist [21] was used. The PREFS checklist was specifically developed on the basis of guidelines for conjoint analyses, patient-reported outcomes and randomised trials to assess studies reporting on factors that could influence the validity of findings. To the best of our knowledge, this is the only checklist available for estimating the quality of preference studies [21]. PREFS assesses five criteria as acceptable (1) or unacceptable (0): the purpose of the study (is the purpose of the study in relation to preferences clearly stated?); respondent sampling (are the responders similar to the non-responders?); explanation of preference assessment methods (are the methods used for assessing preferences clearly explained?); findings reported for the total sample (are all respondents included in the reported findings and analysis of preference results?); and significance testing (are significance tests used to assess the preference results?). The five criteria are summed to indicate an overall quality score ranging from 0 to 5. Studies were scored independently by two reviewers, and disagreements were resolved by consensus in pairs of two researchers (KLC and EMJ, and BFMW and ILH).

3.5 Data Analysis

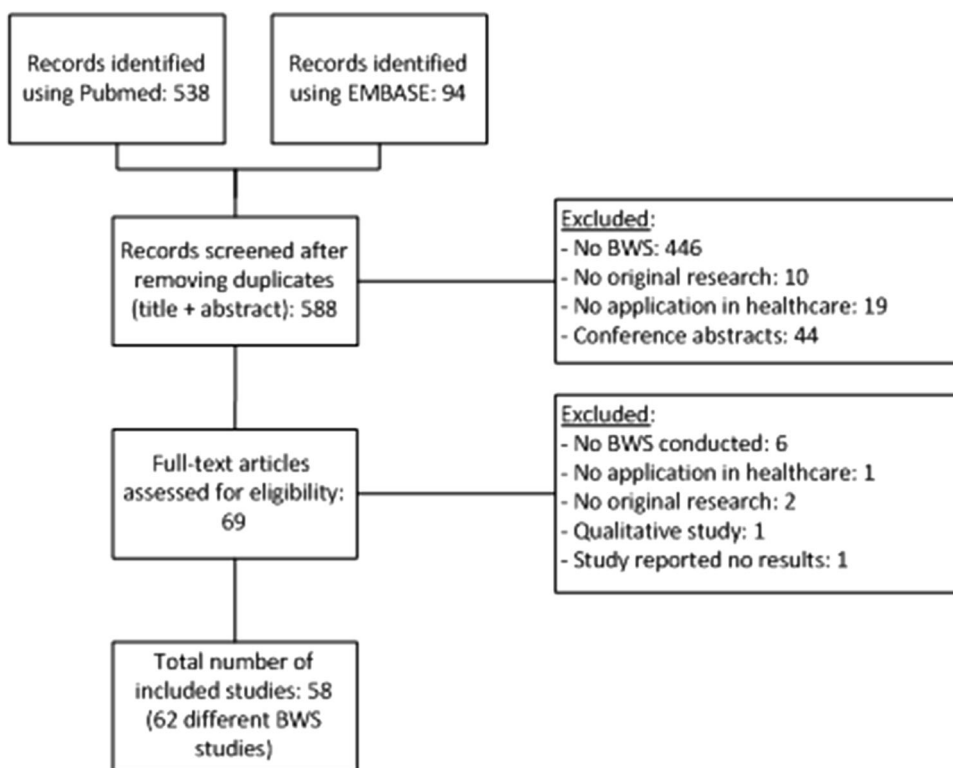
The extracted data were grouped by BWS case type. Additionally, for several items (sample size, number of included factors, maximum number of choice tasks, number of factors per choice task and analytical method), we reported medians and ranges for different time periods (i.e. ≤ 2010 , 2011, 2012, 2013, 2014 and 2015) to investigate potential trends. The extracted data are presented in the tables separately for the three different BWS case types.

4 Results

4.1 Article Inclusion

The initial database searches resulted in 632 articles (538 from Pubmed and 94 from EMBASE). This corresponded to 588 unique studies, of which 69 were eligible for full-text review. Full-text review resulted in exclusion of 11 studies, as six studies did not perform a BWS experiment, one study had no application in health care, two studies were not original research, one study included thinking aloud (qualitative) results only and another study did not report on the results. Figure 2 shows the flow chart of the study identification process.

Fig. 2 Flow chart of the study identification process. BWS best–worst scaling



4.2 General Study Information

A total of 62 BWS studies were identified (58 articles), of which 26 were BWS object case studies, 29 were BWS profile case studies and seven were BWS multi-profile case studies. The results indicated that the numbers of BWS object case and BWS profile case studies applied in health care increased over time (see Fig. 3). The median numbers of participants were 175 for BWS object case, 162 for BWS profile case and 326 for BWS multi-profile case.

Most studies were conducted in the USA (67 %) for BWS object case, and in Australia (36 %) and the UK (32 %) for BWS profile case (see Table 1).

Most studies had health care professionals (BWS object case 27 %, BWS profile case 16 % and BWS multi-profile case 13 %), patients (BWS object case 27 %, BWS profile case 29 % and BWS multi-profile case 13 %) and the general population (BWS object case 27 %, BWS profile case 39 % and BWS multi-profile case 38 %) as the target populations. The main focus of all cases was to explore

Fig. 3 Cumulative numbers of best–worst scaling (BWS) studies by year and by BWS case type

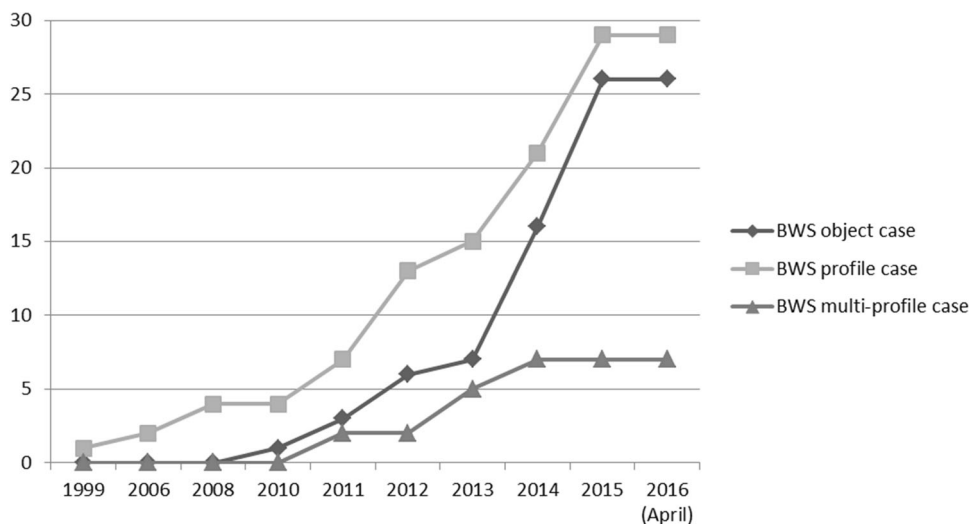


Table 1 Characteristics and quality of best–worst scaling (BWS) studies conducted in health care

Studies ^a	Country/ region	Year	N	Number of factors	Number of choice tasks	Number of factors per choice task	PREFS checklist scores ^b					Total PREFS score
							P	R	E	F	S	
BWS object case												
Louviere and Flynn [32]	Australia	2010	204	15	15	8	1	0	1	1	1	4
Kurkjian et al. (1) [33]	USA	2011	204	11	NR	4	1	0	0	1	0	2
Kurkjian et al. (2) [33]	USA	2011	164	11	NR	4	1	0	0	1	0	2
Gallego et al. [34]	International	2012	120	11	11	5	1	0	1	0	1	3
Marti [35]	Switzerland	2012	376	15	16	11	1	0	1	0	1	4
Mazanov et al. [36]	Australia	2012	168	11	11	5	1	0	1	1	1	3
Silverman et al. [37]	USA	2013	367	39	NR	NR	1	0	0	1	1	3
Cozmuta et al. [38]	USA	2014	118	23	20	5	1	0	1	1	1	4
Ejaz et al. [39]	USA	2014	214	16	16	6	1	0	1	1	1	4
Hauber et al. [40]	USA, Germany	2014	803	10	5	5	1	0	1	1	1	4
Hofstede et al. (1) [41]	Netherlands	2014	246	53	NR	6	1	0	0	0	1	4
Hofstede et al. (2) [41]	Netherlands	2014	155	35	NR	6	1	0	0	0	1	4
Torbica et al. [42]	West Africa	2014	89	11	11	6	1	0	1	1	1	4
van Til et al. [43]	Netherlands, Canada	2014	15	14	12	4	1	1	0	1	0	3
Yuan et al. (1) [44]	USA	2014	273	10	15	3	1	0	1	0	0	2
Yuan et al. (2) [44]	USA	2014	206	9	14	3	1	0	1	0	0	2
Beusterien et al. [45]	USA, UK	2015	245	13	13	4	1	0	1	1	1	4
Fraenkel et al. [46]	USA	2015	162	11	11	5	1	0	1	0	0	2
Hashim et al. [47]	UK	2015	139	13	13	4	1	0	0	1	0	3
Malhotra et al. [48]	Singapore	2015	285	9	12	3	0	0	1	1	0	3
Narurkar et al. [49]	USA	2015	603	14	14	3	1	0	1	0	0	2
Peay et al. [50]	USA	2015	119	16	16	6	1	0	1	1	0	3
Ross et al. [51]	USA	2015	25	16	16	6	1	0	1	1	1	4
Wittenberg et al. [52]	USA	2015	30	11	11	5	1	0	1	NA	NA	2
Yan et al. [53]	USA	2015	110	10	11	5	1	0	1	1	1	4
Yu et al. [54]	USA	2015	182	6	10	3	1	0	1	0	0	2
Median			175	12	13	5						3
BWS profile case												
Szeinbach et al. [11]	USA	1999	33	18	18	6	1	0	0	0	1	2
Coast et al. [55]	UK	2006	96	9	16	4	1	0	1	1	1	4
Coast et al. [56]	UK	2008	255	20	16	5	1	0	1	1	0	3
Flynn et al. [57]	UK	2008	55	10	16	4	1	0	1	0	1	3
Al-Janabi et al. [58]	UK	2011	162	18	18	6	1	0	1	0	1	3
Potoglou et al. [59]	UK	2011	300	34	12	9	1	0	1	1	1	4
Ratcliffe et al. [60]	Australia	2011	16	45	5	9	1	0	1	1	0	3
Knox et al. [61]	Australia	2012	362	42	16	7	1	1	1	1	1	4
Knox et al. [61]	Australia	2012	362	42	16	7	1	1	1	1	1	4
Molassiotis et al. [62]	UK	2012	87	8	16	4	1	0	1	1	1	4
Najafzadeh et al. [63]	Canada	2012	197	16	16	6	1	0	1	1	1	4
Netten et al. [64]	UK	2012	1296	32	8	8	1	1	1	0	1	4
Ratcliffe et al. [65]	Australia	2012	590	45	10	9	1	0	1	0	1	3
Severin et al. [66]	Europe	2013	26	13	12	6	1	0	1	1	1	4
Yoo and Dorion (1) [67]	Australia	2013	526	26	8	12	1	1	1	1	1	5
Damery et al. [68]	UK	2014	132	12	9	4	1	0	1	1	1	4
Hollin et al. [69]	USA	2014	119	18	18	6	1	0	1	1	1	4

Table 1 continued

Studies ^a	Country/ region	Year	<i>N</i>	Number of factors	Number of choice tasks	Number of factors per choice task	PREFS checklist scores ^b					Total PREFS score
							<i>P</i>	<i>R</i>	<i>E</i>	<i>F</i>	<i>S</i>	
Peay et al. [70]	USA	2014	119	18	18	6	1	0	1	0	1	3
Ratcliffe [71]	Australia	2014	24	23	6	7	1	0	1	1	0	3
Ungar et al. [72]	Canada	2014	101	16	16	6	1	0	1	1	1	4
Whitty et al. [73]	Australia	2014	930	23	7	7	1	0	1	1	1	4
dosReis et al. [74]	USA	2015	37	21	18	7	1	0	1	1	1	4
Flynn et al. [75]	UK	2015	413	20	16	5	1	0	1	0	1	3
Franco et al. [76]	Australia	2015	220	45	10	9	1	0	1	0	1	3
Gendall et al. [77]	New Zealand	2015	534	15	10	3	1	0	1	1	0	4
Jones et al. [78]	Australia	2015	31	10	32	4	1	1	1	1	1	5
O'Hara et al. [79]	South Africa	2015	125	15	12	5	1	0	1	1	1	4
Ratcliffe et al. [80]	Australia	2015	1190	45	10	9	1	0	1	0	0	2
Tsao et al. [81]	Canada	2015	819	13	13	4	0	0	1	1	0	3
Median			162	18	16	6						4
BWS multi-profile case												
Brown et al. [82]	USA	2011	53	36	12	12	1	0	1	1	1	4
Hoek et al. [83]	New Zealand	2011	292	13	13	4	1	0	1	1	1	4
Cameron et al. [84]	Thailand	2013	326	14	1	7	1	0	1	0	1	4
Lancsar et al. [22]	Canada	2013	898	10	16	5	1	0	1	0	1	3
Yoo and Doiron (2) [67]	Australia	2013	526	26	8	12	1	1	1	1	1	5
Maubach et al. [85]	New Zealand	2014	768	14	9	4	1	0	1	1	1	4
Xie et al. [86]	Canada	2014	100	25	11	5	1	0	1	1	1	4
Median			326	14	11	5						4

NA not applicable, NR not reported

^a Numbers in parentheses represent 2 BWS studies described in the same article

^b The PREFS checklist criteria are Purpose, Respondents, Explanation, Findings, Significance

patient and consumer experience factors (BWS object case 32 %, BWS profile case 45 % and BWS multi-profile case 40 %; see Table 2). A table including a detailed outline of the study characteristics can be found in Electronic Supplementary Material 2.

4.3 Design and Choice Sets

Looking at current practice regarding the designs and choice sets used in BWS studies, most BWS studies selected factors (attributes and levels) by using literature reviews (BWS object case 33 %, BWS profile case 35 % and BWS multi-profile case 50 %) and/or qualitative methods, such as interviews and focus groups (BWS object case 55 %, BWS profile case 59 % and BWS multi-profile case 25 %). For BWS object case, the median number of factors was 12, while the average number of factors in BWS profile case (defined as the number of levels included in the experiment) was 18. For BWS multi-profile case, the median number of factors was 14. The median numbers of choice sets were 13 for BWS object case, 16 for BWS

profile case and 11 for BWS multi-profile case. Almost all studies asked respondents to identify only one best and one worst option. Only one study (BWS multi-profile case) asked respondents to identify two best and two worst choices for each choice set [22].

In most studies, the survey was administrated in a self-completion format (BWS object case 74 %, BWS profile case 72 % and BWS multi-profile case 71 %), and many studies used an online platform (BWS object case 48 %, BWS profile case 41 % and BWS multi-profile case 29 %). The most commonly used designs by far were the balanced incomplete block design for BWS object case (54 %) and the orthogonal main effects design for BWS profile case (66 %) and BWS multi-profile case (43 %). However, some studies did not report on their designs (BWS object case 15 %, BWS profile case 7 % and BWS multi-profile case 14 %).

4.4 Estimation Procedure and Validity

As shown in Table 2, BWS object case studies most often employed hierarchical Bayes estimation (21 %) and simple

Table 2 Characteristics of best–worst scaling (BWS) studies conducted in health care

Item	Category	BWS object case		BWS profile case		BWS multi-profile case	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Target population ^a	Health care professionals	7	27	5	16	1	13
	Patients	7	27	9	29	1	13
	General population	7	27	12	39	3	38
	(Informal) caregivers	3	12	4	13	1	13
	Policy makers	1	4	0	0	0	0
	Other stakeholders	1	4	1	3	2	25
Area of application ^a	Valuing health outcomes	10	32	13	45	4	40
	Investigating trade-offs between health outcomes and patient or consumer experience factors	8	26	4	14	0	0
	Estimating utility weights within the quality-adjusted life-year framework	2	6	1	3	4	40
	Job choices	0	0	3	10	1	10
	Developing a priority-setting mechanism	0	0	1	3	1	10
	Health professionals' preferences for treatment and screening options	4	13	4	14	0	0
	Other	6	19	1	3	0	0
BWS design	Orthogonal main effects	3	12	19	66	3	43
	Latin square (balanced order and pairing)	2	8	0	0	0	0
	Balanced incomplete block	14	54	2	7	1	14
	Full factorial	2	8	2	7	0	0
	Bayesian efficient	1	4	2	7	1	14
	Fractional other	0	0	2	7	1	14
	NR	4	15	2	7	1	14
Domains of BWS factors ^a	Money	8	12	9	13	4	27
	Time	6	9	7	10	1	7
	Risk	7	11	6	8	2	13
	Health care	16	24	17	24	3	20
	Health status	15	23	16	22	2	13
	Other	14	21	17	24	3	20
Analytical method ^a	Hierarchical Bayes	7	21	1	3	0	0
	Simple summary statistics (best-minus-worst summary statistics)	11	32	8	24	0	0
	MNL model	4	12	16	47	5	63
	Weighted least squares	2	6	5	15	1	0
	Latent class analysis	3	9	1	3	0	13
	Max diff scaling	2	6	0	0	2	0
	Ordered logit	3	9	0	0	0	25
	Random parameter logit model	0	0	1	3	0	0
	Qualitative; thinking aloud procedure	2	6	0	0	0	0
	NR	7	21	1	3	0	0
Heterogeneity	No heterogeneity	17	68	19	66	4	14
	Latent class	2	8	6	21	0	0
	Random parameter	4	16	4	14	3	10
	NR	2	8	0	0	0	0

Table 2 continued

Item	Category	BWS object case		BWS profile case		BWS multi-profile case	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Software used for analyses of BWS ^a	Sawtooth software	7	27	2	7	0	0
	SAS	2	8	3	10	2	29
	Stata	2	8	9	31	1	14
	Nlogit	0	0	3	10	0	0
	SPSS	2	8	0	0	0	0
	Latent Gold Choice	0	0	2	7	0	0
	NA/NR	13	50	10	34	4	57

Max diff maximum difference scaling, *MNL* multinomial logistic regression, *NA* not applicable, *NR* not reported

^a More than one category per BWS study was possible

summary statistics (32 %). BWS profile case and BWS multi-profile case studies used the multinomial logistic regression (MNL) model (in 47 % and 63 % of the studies, respectively). Several studies did not report on the software used for analyses (BWS object case 50 %, BWS profile case 34 % and BWS multi-profile case 57 %). Of those that did report on the software used for analysis, Sawtooth software was most frequently used for BWS object case (27 %), while Stata software was most frequently used in analyses of BWS profile case (31 %), and SAS was most frequently reported in BWS multi-profile case studies (29 %). Most studies did not take into account heterogeneity (BWS object case 68 %, BWS profile case 66 % and BWS multi-profile case 57 %). To validate the factors, a pilot study was used for 15 BWS object case studies (58 %), 16 BWS profile case studies (55 %) and three BWS multi-profile case studies (43 %). In addition, a large majority of BWS studies conducted no validity checks (BWS object case 81 %, BWS profile case 76 % and BWS multi-profile case 71 %).

4.5 Trend Analysis

Table 3 depicts the trend analysis, which indicated a decreasing median sample size since 2013 for BWS object case (from $N = 367$ in articles published in 2013 to $N = 151$ in articles published in 2015). For BWS profile case and BWS multi-profile case, the trends were not so clear, which is also depicted by the wide ranges around the medians. Moreover, there appeared to be a decrease in the median number of factors included in BWS object case studies since 2013. For BWS profile case, a similar trend since 2011 was found. For BWS multi-profile case, the trend was not so clear. For BWS profile case, a decrease in the median number of factors per choice task since 2013 was found. No trends could be found regarding the median numbers of choice tasks

and the median numbers of factors per choice task in BWS studies. However, looking at the maximum values of the numbers of choice tasks in BWS profile case studies, there seemed to be a slight increase from 2010 onwards. The trends in the analytical methods are depicted in Fig. 4a–c. There was a steep increase in the use of best-minus-worst counting and hierarchical Bayes estimation in BWS object case studies (see Fig. 4a). In BWS profile case studies, a steep increase was observed in the use of MNL models (see Fig. 4b). With regard to BWS multi-profile case studies, the trends were not clear (see Fig. 4c).

4.6 Quality of Best–Worst Scaling Studies

The scores on the PREFS checklist used to assess the quality of current BWS studies are included in Table 1. We found that 98 % of the BWS studies had a preference-related purpose clearly stated. However, most studies did not report whether the responders were similar to the non-responders, with 56 studies (91 %) categorised as unacceptable with regard to this criterion. We found that 54 studies (86 %) clearly explained the methods used to assess preferences. Most studies included all respondents in the analyses of preference results and reported their findings (65 %). Furthermore, 44 studies (71 %) used significance tests to assess the preference results. Regarding the total PREFS scores, half of the studies scored 4 or higher (56.4 %), of which only three studies achieved a score of 5.

5 Discussion

This study reviewed current practice in the use and conduct of BWS within health care. The use of BWS—especially BWS object case and BWS profile case—has drastically increased in health care, with a steep increase in the last 2

Table 3 Trend analysis

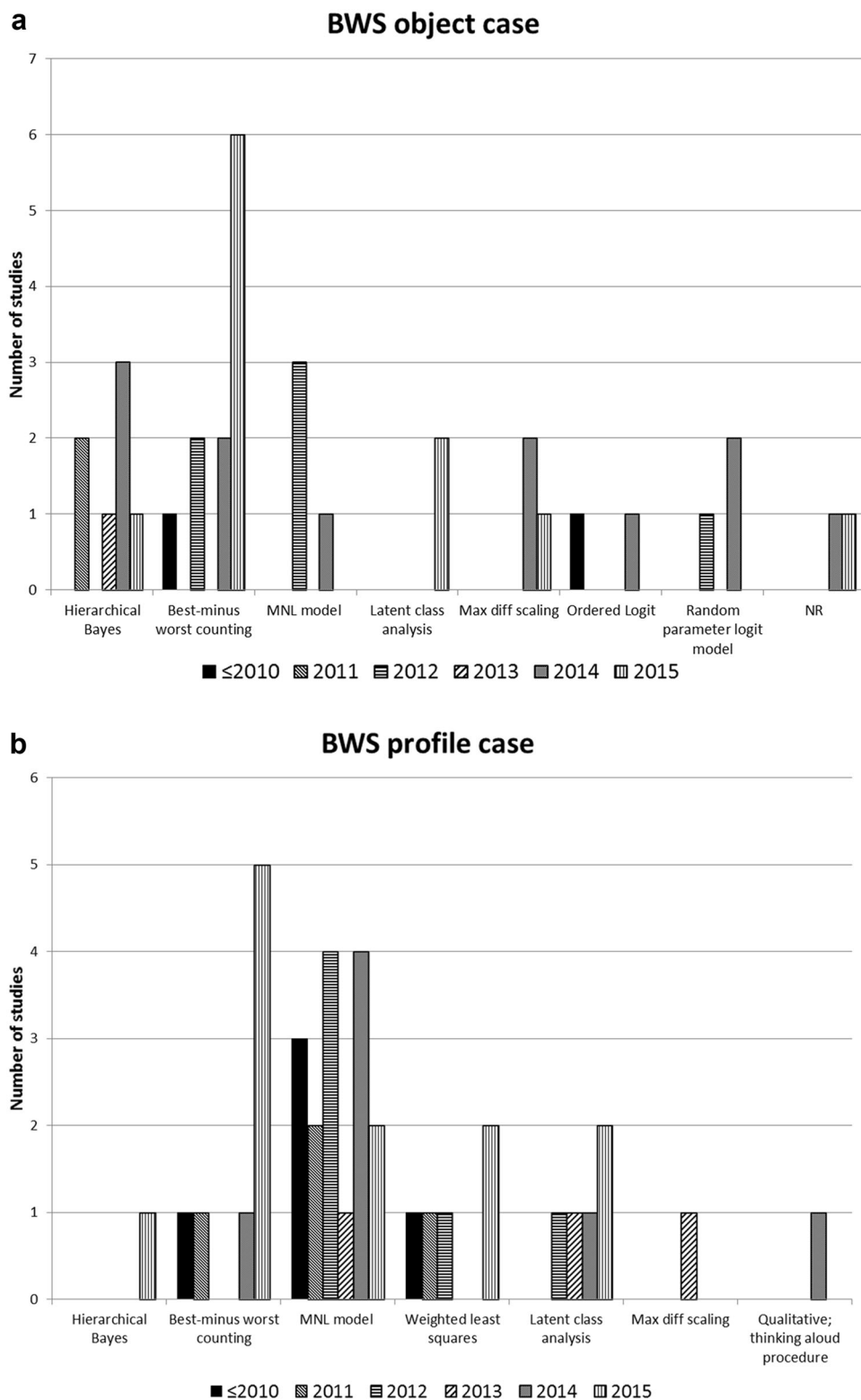
	BWS object case				BWS profile case				BWS multi-profile case			
	Median	Min	Max	<i>N</i>	Median	Min	Max	<i>N</i>	Median	Min	Max	<i>N</i>
Number of included factors												
≤2010	15	15	15	1	14	9	20	4	0	0	0	0
2011	11	11	11	2	34	18	45	3	24.5	13	36	2
2012	11	11	15	3	37	8	45	6	0	0	0	0
2013	39	39	39	1	19.5	13	26	2	14	10	26	3
2014	14	9	53	9	18	12	23	6	19.5	14	25	2
2015	12	6	16	10	17.5	10	45	8	0	0	0	0
Number of factors per choice task												
≤2010	8	8	8	1	4.5	4	6	4	0	0	0	0
2011	4	4	4	2	9	6	9	3	8	4	12	2
2012	5	5	11	3	7	4	9	6	0	0	0	0
2013	NR	NR	NR	1	9	6	12	2	7	5	12	3
2014	5	3	6	9	6	4	7	6	4.5	4	5	2
2015	4.5	3	6	10	5	3	9	8	0	0	0	0
Number of choice tasks												
≤2010	15	15	15	1	16	16	18	4	0	0	0	0
2011	NR	NR	NR	2	12	5	18	3	12.5	12	13	2
2012	11	11	16	3	16	8	16	6	0	0	0	0
2013	NR	NR	NR	1	10	8	12	2	8	1	16	3
2014	14	5	20	9	12.5	6	18	6	10	9	11	2
2015	12.5	10	16	10	12.5	10	32	8	0	0	0	0
Sample size												
≤2010	204	204	204	1	75.5	33	255	4	0	0	0	0
2011	184	164	204	2	162	16	300	3	172.5	53	292	2
2012	168	120	376	3	362	87	1296	6	0	0	0	0
2013	367	367	367	1	276	26	526	2	526	326	898	3
2014	206	15	803	9	119	24	930	6	434	100	768	2
2015	150.5	25	603	10	316.5	31	1190	8	0	0	0	0

BWS best–worst scaling, *Max* maximum, *Min* minimum, *NR* not reported

years in a wide range of applications (e.g. valuing health outcomes and investigating preferences between health outcomes and patient or consumer experience factors) and populations (e.g. health care professionals, patients, the general population and caregivers). This increase in published BWS studies may indicate increasing awareness of the advantages of the BWS method to elicit preferences in health care. It is apparent that BWS, especially BWS object case, is capable of incorporating a large set of items or factors to elicit preferences in comparison with DCE. In DCE, the average number of included factors was estimated to be 4–6 [14], compared with an average of 17 (median 12) factors for BWS object case. BWS object case can be especially useful if one is interested in a list of items that is difficult to reduce to a number acceptable for other conjoint analyses, such as DCE. Additionally, in comparison with DCE, BWS (especially BWS object case and BWS profile case) may be less cognitively burdensome,

which can be valuable in health care settings where preferences are often elicited from vulnerable patient groups [9]. For example, it has been stated that best–worst tasks take advantage of a person’s propensity to identify and respond more consistently to extreme options, and BWS tasks seem easier to understand for people [9, 23]. In comparison with the results from a DCE, BWS may give different findings especially because choosing a ‘worst’ option may be a different psychological process. The literature has shown mixed results as to whether similar patterns arise between the two methods, and further comparison would be needed [23–25]. Hence, one should be careful when directly comparing the results of DCE and BWS. Despite the observed increase in the use of BWS within health care, the number of published BWS studies still lags behind the number of DCE studies. Clark et al. [14] identified 179 DCE studies published between 1990 and 2012, which is more than three times the total of 58

Fig. 4 Analytical methods used per year: **a** best–worst scaling (BWS) object case, **b** BWS profile case and **c** BWS multi-profile case. *Max diff* maximum difference scaling, *MNL* multinomial logistic regression, *NR* not reported

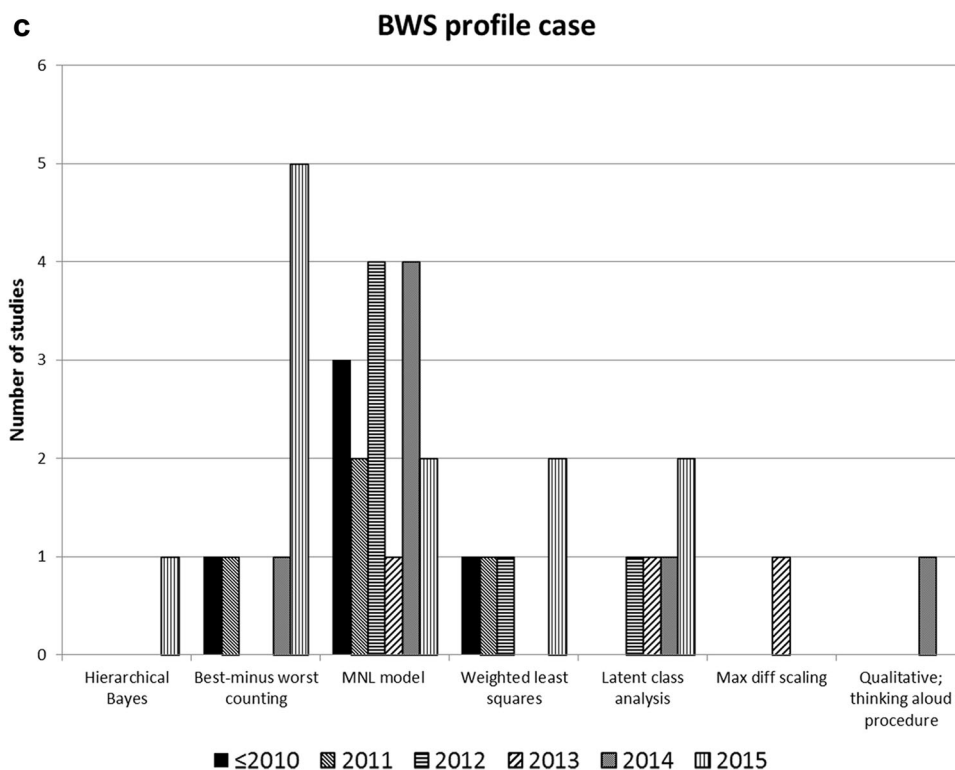


published articles identified in this paper up until April 2016.

In comparison with BWS object case and BWS profile case, a notable difference in the use of BWS multi-profile case within health care was observed. From this review, it

is apparent that there is limited application of BWS multi-profile case, which may have several explanations. One explanation is that researchers have difficulty differentiating BWS multi-profile case from DCE or fail to see added value in using BWS multi-profile case over DCE. There are

Fig. 4 continued



only minor differences between both methods, which may explain why BWS multi-profile case is often referred to as ‘BWS DCE’ [26, 27]. Although BWS multi-profile case has been stated to offer distinct advantages, such as providing richer information than methods based on a ‘pick one’ approach [13, 27] (e.g. resulting in a smaller sample size being needed), the need to elicit the best and worst alternatives has the limitation of comparing a minimum of three profiles per choice task.

In our review, several trends regarding changes in key aspects of BWS were identified. First, regarding the sample size for BWS object case, we observed use of smaller sample sizes since 2013. The literature provides no guidance as to the minimal sample size for BWS application. Future research should focus on developing sample size calculations based on the desired statistical power, similar to DCE [28]. Second, the median numbers of factors included in BWS object case and BWS profile case studies have decreased substantially since 2013 and 2011, respectively. For DCEs, the literature indicates no clear trends regarding the number of factors used [14]. Yet, this study revealed a trend towards a decreasing number of factors being included in BWS studies. Third, when the type of analytical method is assessed, relatively less complicated methods (e.g. best–worst counting) have become increasingly popular, especially in BWS object case studies. In contrast with DCE, where more advanced methods (such as models that take heterogeneity into account, e.g. a

latent class or random parameter model) are used [14], BWS can be analysed more easily, making the method attractive to a wide range of researchers. Details regarding the analytical methods underlying BWS can be found in Mühlbacher et al. [17, 18].

The quality of BWS studies in general was acceptable, but there is room for improvement, as still nearly half of the studies (44 %) scored either 2 or 3 on a 5-point scale. Of all studies, 35 % did not include all responders in the final analyses. However, many of the studies that did not include all respondents did so for a specific reason (e.g. failed comprehension tests or missing data). Most BWS studies (90 %) did not report whether the responders were similar to non-responders. This resulted in a lack of transparency, which may have introduced a non-response bias. Future BWS studies are thus recommended to track non-responders on some (demographic) variables and perform comparative analyses if possible. However, because of the increased popularity of online panels, this may not always be feasible. In addition, PREFS scores of 0 should be interpreted with caution, as low PREFS scores are not synonymous with low quality. For instance, not using all respondents for reasons of internal validity (e.g. failure on a dominance test) or missing values may be an appropriate choice.

Our study had some potential limitations. First, as the aim of this study was to identify current practice of BWS within health care, the study findings and trends may not be

applicable/generalisable to BWS studies outside the health care setting. Second, as the number of BWS multi-profile case studies is still limited, statements made regarding BWS multi-profile case in this paper should be treated with caution and are provided only to give some initial insights into the conduct of BWS multi-profile case studies within health care. Third, we identified some discrepancies between reviewers in terms of our data collection sheet and how some fields were being interpreted (e.g. validity checks and heterogeneity). However, we resolved those discrepancies in discussions among the reviewers throughout the study. Fourth, this study looked only at BWS as defined by Louviere et al. [19], including both best and worst questions. Recently, new alternatives/variations for best–worst tasks have been published, such as best–best scaling or best–only scaling [29–31]. To enhance comparability and clarity, studies of this variety were not included in this review. Lastly, this study included only published articles and did not search among conference proceedings.

6 Conclusion

The use of BWS object case and BWS profile case has drastically increased in health care, especially in the last 2 years. The application of BWS multi-profile case is, however, still very limited. In recent studies, trends towards decreasing sample sizes and decreasing numbers of factors in BWS object case—as well as use of less complicated analytical methods—have been observed, enhancing the usefulness of the BWS method for wider health care problems and making it accessible to more researchers. Next to that, the quality of BWS studies is generally acceptable, except with regard to indicating whether responders are similar to non-responders. This implies the need to increase the transparency of BWS studies and check for non-response bias when possible. Furthermore, this study may serve as a vantage point reflecting state-of-the-art practices in BWS studies.

Acknowledgments Kei Long Cheung, Ben Wijnen and Mickael Hilgsmann designed and planned the study. Kei Long Cheung and Ben Wijnen screened the titles, abstracts and full texts of the studies considered for inclusion in this review, and produced the first draft of the manuscript. Kei Long Cheung, Ben Wijnen, Ilene Hollin and Ellen Janssen extracted the data and appraised the quality of the studies. Different versions of the manuscript were reviewed and conceptualised by all authors. All authors read and approved the final manuscript. No funding was received for conduct of this study and/or preparation of this manuscript.

Compliance with Ethical Standards

Conflict of interest All authors (Kei Long Cheung, Ben Wijnen, Ilene Hollin, Ellen Janssen, John Bridges, Silvia Evers and Mickael Hilgsmann) declare that they have no competing interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bridges JF, Jones C. Patient-based health technology assessment: a vision of the future. *Int J Technol Assess Health Care*. 2007;23(1):30–5.
2. 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.
3. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403–13.
4. Janz NK, Wren PA, Copeland LA, et al. Patient–physician concordance: preferences, perceptions, and factors influencing the breast cancer surgical decision. *J Clin Oncol*. 2004;22(15):3091–8.
5. Facey K, Boivin A, Gracia J, et al. Patients’ perspectives in health technology assessment: a route to robust evidence and fair deliberation. *Int J Technol Assess Health Care*. 2010;26(3):334–40.
6. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403–13.
7. Green PE, Srinivasan V. Conjoint analysis in consumer research: issues and outlook. *J Consum Res*. 1978;5:103–23.
8. Bridges JF, Kinter ET, Kidane L, et al. Things are looking up since we started listening to patients. *Patient*. 2008;1(4):273–82.
9. Marley AA, Louviere JJ. Some probabilistic models of best, worst, and best–worst choices. *J Math Psychol*. 2005;49(6):464–80.
10. Finn A, Louviere JJ. Determining the appropriate response to evidence of public concern: the case of food safety. *J Public Policy Mark*. 1992;11(2):12–25.
11. Szeinbach SL, Barnes JH, McGhan WF, et al. Using conjoint analysis to evaluate health state preferences. *Drug Inf J*. 1999;33(3):849–58.
12. McIntosh E, Louviere JJ. Separating weight and scale value: an exploration of best-attribute scaling in health economics. *Odense: Health Econ Study Group Meet*; 2002.
13. Flynn TN, Louviere JJ, Peters TJ, et al. Best–worst scaling: what it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171–89.
14. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 2014;32(9):883–902.
15. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145–72.
16. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. *Exp Rev Pharmacoecon Outcomes Res*. 2010;10(3):259–67.
17. Mühlbacher AC, Kaczynski A, Zweifel P, et al. Experimental measurement of preferences in health and healthcare using best–worst scaling: an overview. *Health Econ Rev*. 2016;6(1):1–14.

18. Mühlbacher AC, Zweifel P, Kaczynski A, et al. Experimental measurement of preferences in health care using best–worst scaling (BWS): theoretical and statistical issues. *Health Econ Rev.* 2016;6(1):1–12.
19. Louviere JJ, Flynn TN, Marley A. Best–worst scaling: theory, methods and applications. Cambridge: Cambridge University Press; 2015.
20. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* 2012;21(2):145–72.
21. Joy SM, Little E, Maruthur NM, et al. Patient preferences for the treatment of type 2 diabetes: a scoping review. *Pharmacoeconomics.* 2013;31(10):877–92.
22. Lancsar E, Louviere J, Donaldson C, et al. Best worst discrete choice experiments in health: methods and an application. *Soc Sci Med.* 2013;76(1):74–82.
23. 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):1717–27.
24. Severin F, Schmidtke J, Mühlbacher A, et al. 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):1202–8.
25. Louviere JJ, Islam T. A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *J Bus Res.* 2008;61(9):903–11.
26. Gallego G, Dew A, Lincoln M, 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. *Hum Resour Health.* 2015;13(1):53.
27. Lancsar E, Louviere J, Donaldson C, et al. Best worst discrete choice experiments in health: methods and an application. *Soc Sci Med.* 2013;76:74–82.
28. de Bekker-Grob EW, Donkers B, Jonker MF, et al. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient.* 2015;8(5):373–84.
29. Ghijben P, Lancsar E, Zavarsek S. Preferences for oral anticoagulants in atrial fibrillation: a best–best discrete choice experiment. *Pharmacoeconomics.* 2014;32(11):1115–27.
30. Meyfroidt S, Hulscher M, De Cock D, et al. A maximum difference scaling survey of barriers to intensive combination treatment strategies with glucocorticoids in early rheumatoid arthritis. *Clin Rheumatol.* 2015;34(5):861–9.
31. Van Hulst L, Kievit W, Van Bommel R, et al. Rheumatoid arthritis patients and rheumatologists approach the decision to escalate care differently: results of a maximum difference scaling experiment. *Arthritis Care Res.* 2011;63(10):1407–14.
32. Louviere JJ, Flynn TN. Using best–worst scaling choice experiments to measure public perceptions and preferences for health-care reform in Australia. *Patient.* 2010;3(4):275–83.
33. Kurkjian TJ, Kenkel JM, Sykes JM, et al. Impact of the current economy on facial aesthetic surgery. *Aesthet Surg J.* 2011;31(7):770–4.
34. Gallego G, Bridges JFP, Flynn T, et al. Using best–worst scaling in horizon scanning for hepatocellular carcinoma technologies. *Int J Technol Assess Health Care.* 2012;28(3):339–46.
35. Marti J. A best–worst scaling survey of adolescents’ level of concern for health and non-health consequences of smoking. *Soc Sci Med.* 2012;75(1):87–97.
36. Mazanov J, Huybers T, Connor J. Prioritising health in antidoping: what Australians think. *J Sci Med Sport.* 2012;15(5):381–5.
37. 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;24(7):2067–77.
38. Cozmata R, Merkel PA, Wahl E, et al. Variability of the impact of adverse events on physicians’ decision making. *BMC Med Inform Decis Mak.* 2014;14:86.
39. Ejaz A, Spolverato G, Bridges JF, et al. Choosing a cancer surgeon: analyzing factors in patient decision making using a best–worst scaling methodology. *Ann Surg Oncol.* 2014;21(12):3732–8.
40. Hauber AB, Mohamed AF, Johnson FR, et al. Understanding the relative importance of preserving functional abilities in Alzheimer’s disease in the United States and Germany. *Qual Life Res.* 2014;23(6):1813–21.
41. Hofstede SN, van Bodegom-Vos L, Wentink MM, et al. Most important factors for the implementation of shared decision making in sciatica care: ranking among professionals and patients. *PLoS ONE.* 2014;9(4):e94176.
42. Torbica A, De Allegri M, Belemsaga D, et al. What criteria guide national entrepreneurs’ policy decisions on user fee removal for maternal health care services? Use of a best–worst scaling choice experiment in West Africa. *J Health Serv Res Policy.* 2014;19(4):208–15.
43. van Til J, Groothuis-Oudshoorn C, Lieferink M, et al. Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework. *Cost Eff Resour Alloc.* 2014;12(1):22.
44. Yuan Z, Levitan B, Burton P, et al. Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives. *Curr Med Res Opin.* 2014;30(9):1733–41.
45. Beusterien K, Kennelly MJ, Bridges JF, et al. Use of best–worst scaling to assess patient perceptions of treatments for refractory overactive bladder. *Neurourol Urodyn.* 2015. doi:10.1002/nau.22876 (Epub 2015 Sep 14).
46. Fraenkel L, Lim J, Garcia-Tsao G, et al. Variation in treatment priorities for chronic hepatitis C: a latent class analysis. *Patient.* 2016;9(3):241–9.
47. Hashim H, Beusterien K, Bridges JF, et al. Patient preferences for treating refractory overactive bladder in the UK. *Int Urol Nephrol.* 2015;47(10):1619–27.
48. Malhotra C, Chan N, Zhou J, et al. Variation in physician recommendations, knowledge and perceived roles regarding provision of end-of-life care. *BMC Palliat Care.* 2015;14(1):52.
49. Narurkar V, Shamban A, Sissins P, et al. Facial treatment preferences in aesthetically aware women. *Dermatol Surg.* 2015;41(Suppl 1):S153–60.
50. Peay HL, Hollin I, Bridges J. Prioritizing parental worry associated with Duchenne muscular dystrophy using best–worst scaling. *J Genet Couns.* 2016;25(2):305–13.
51. Ross M, Bridges JFP, Ng X, et al. A best–worst scaling experiment to prioritize caregiver concerns about ADHD medication for children. *Psychiatr Serv.* 2015;66(2):208–11.
52. Wittenberg E, Bharel M, Saada A, et al. Measuring the preferences of homeless women for cervical cancer screening interventions: development of a best–worst scaling survey. *Patient.* 2015;8(5):455–67. doi:10.1007/s40271-014-0110-z (Epub 2015).
53. Yan K, Bridges JF, Augustin S, et al. Factors impacting physicians’ decisions to prevent variceal hemorrhage. *BMC Gastroenterol.* 2015;15:55.
54. Yu T, Holbrook JT, Thorne JE, et al. Outcome preferences in patients with noninfectious uveitis: results of a best–worst scaling study. *Invest Ophthalmol Vis Sci.* 2015;56(11):6864–72.
55. Coast J, Flynn TN, Salisbury C, et al. Maximising responses to discrete choice experiments: a randomised trial. *Appl Health Econ Health Policy.* 2006;5(4):249–60.

56. Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP Capability Index for older people. *Soc Sci Med*. 2008;67(5): 874–82.
57. Flynn TN, Louviere JJ, Peters TJ, et al. Estimating preferences for a dermatology consultation using best–worst scaling: comparison of various methods of analysis. *BMC Med Res Methodol*. 2008;8:76.
58. Al-Janabi H, Flynn TN, Coast J. Estimation of a preference-based carer experience scale. *Med Decis Making*. 2011;31(3):458–68.
59. 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):1717–27.
60. Ratcliffe J, Couzner L, Flynn T, et al. Valuing child health utility 9D health states with a young adolescent sample: a feasibility study to compare best–worst scaling discrete-choice experiment, standard gamble and time trade-off methods. *Appl Health Econ Health Policy*. 2011;9(1):15–27.
61. Knox SA, Viney RC, Street DJ, et al. What's good and bad about contraceptive products? A best–worst attribute experiment comparing the values of women consumers and GPs. *Pharmacoeconomics*. 2012;30(12):1187–202.
62. Molassiotis A, Emsley R, Ashcroft D, et al. Applying best–worst scaling methodology to establish delivery preferences of a symptom supportive care intervention in patients with lung cancer. *Lung cancer (Amsterdam, Netherlands)*. 2012;77(1): 199–204.
63. Najafzadeh M, Lynd LD, Davis JC, et al. Barriers to integrating personalized medicine into clinical practice: a best–worst scaling choice experiment. *Genet Med*. 2012;14(5):520–6.
64. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16(16):1–165.
65. Ratcliffe J, Flynn T, Terlich F, et al. Developing adolescent-specific health state values for economic evaluation: an application of profile case best–worst scaling to the child health utility 9D. *Pharmacoeconomics*. 2012;30(8):713–27.
66. Severin F, Schmidtke J, Mühlbacher A, et al. 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):1202–8.
67. Yoo HI, Doiron D. The use of alternative preference elicitation methods in complex discrete choice experiments. *J Health Econ*. 2013;32(6):1166–79.
68. Damery S, Biswas M, Billingham L, et al. Patient preferences for clinical follow-up after primary treatment for soft tissue sarcoma: a cross-sectional survey and discrete choice experiment. *Eur J Surg Oncol*. 2014;40(12):1655–61.
69. 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;8(1):19–27.
70. Peay HL, Hollin I, Fischer R, et al. A community-engaged approach to quantifying caregiver preferences for the benefits and risks of emerging therapies for Duchenne muscular dystrophy. *Clin therapeutics*. 2014;36(5):624–37.
71. 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.
72. Ungar WJ, Hadioonzadeh A, Najafzadeh M, et al. Quantifying preferences for asthma control in parents and adolescents using best–worst scaling. *Respir Med*. 2014;108(6):842–51.
73. Whitty JA, Ratcliffe J, Chen G, et al. Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best–worst scaling methods. *Med Decis Making*. 2014;34(5):638–54.
74. dosReis S, Ng X, Frosch E, et al. Using best–worst scaling to measure caregiver preferences for managing their child's ADHD: a pilot study. *Patient*. 2015;8(5):423–31.
75. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP—a capability instrument. Estimation of a UK general population tariff. *Health Econ*. 2015;24(3):258–69.
76. Franco MR, Howard K, Sherrington C, et al. Eliciting older people's preferences for exercise programs: a best–worst scaling choice experiment. *J Physiother*. 2015;61(1):34–41.
77. Gendall P, Hoek J, Taylor R, et al. Should support for obesity interventions or perceptions of their perceived effectiveness shape policy? *Aust N Z J Public Health*. 2015;39(2):172–6.
78. Jones LG, Hawkins GE, Brown SD. Using best–worst scaling to improve psychological service delivery: an innovative tool for psychologists in organized care settings. *Psychol Serv*. 2015;12(1): 20–7.
79. O'Hara NN, Roy L, O'Hara LM, et al. Healthcare worker preferences for active tuberculosis case finding programs in South Africa: a best–worst scaling choice experiment. *PLoS ONE*. 2015;10(7):e0133304.
80. Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. 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 Econ*. 2016;25(4):486–96. doi:10.1002/hec.3165.
81. Tsao NW, Khakban A, Gastonguay L, et al. Perceptions of British Columbia residents and their willingness to pay for medication management services provided by pharmacists. *Can Pharm J*. 2015;148(5):263–73.
82. Brown TM, Pashos CL, Joshi AV, et al. The perspective of patients with haemophilia with inhibitors and their care givers: preferences for treatment characteristics. *Haemophilia*. 2011;17(3): 476–82.
83. Hoek J, Wong C, Gendall P, et al. Effects of dissuasive packaging on young adult smokers. *Tob Control*. 2011;20(3):183–8.
84. Cameron MP, Newman PA, Rongprakhon S, et al. The marginal willingness-to-pay for attributes of a hypothetical HIV vaccine. *Vaccine*. 2013;31(36):3712–7.
85. Maubach N, Hoek J, Mather D. Interpretive front-of-pack nutrition labels: comparing competing recommendations. *Appetite*. 2014;82:67–77.
86. Xie F, Pullenayegum E, Gaebel K, et al. Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best–worst scaling? *Eur J Health Econ*. 2014;15(3):281–8.