Journal of Choice Modelling 41 (2021) 100325

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Case 2 best-worst scaling: For good or for bad but not for both

V. Soekhai [a,b,c,*], B. Donkers [a,d], B. Levitan [e], E.W. de Bekker-Grob [a,b]

[a] Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, the Netherlands
[b] Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[c] Erasmus MC-Erasmus University Medical Center Rotterdam, Rotterdam, Netherlands
[d] Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands
[e] Janssen Research & Development, Titusville, NJ, United States

ARTICLE INFO

Keywords:
Best-worst scaling
Dominance
Parameter estimation
Preferences

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.

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).[1] The object case (Fig. 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 (Fig. 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

* Corresponding author. Erasmus Choice Modelling Centre (ECMC), Erasmus School of Health Policy & Management (ESHPM), Erasmus MC - Erasmus University Medical Center Rotterdam (EMC), Erasmus University Rotterdam (EUR), P.O. Box 1738, 3000, DR, Rotterdam, the Netherlands.
E-mail addresses: soekhai@eshpm.eur.nl, v.r.soekhai@gmail.com (V. Soekhai).

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

https://doi.org/10.1016/j.jocm.2021.100325
Received 18 November 2020; Received in revised form 17 September 2021; Accepted 24 September 2021
Available online 1 October 2021

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

Fig. 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).

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 (Fig. 1c) includes two or more profiles where individuals choose their "best" and "worst" profiles. The multi-profile 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.

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.

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.

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).[2] This results in estimation failures and as a consequence such studies are unlikely to result in publications.

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.

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 Fig. 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 Fig. 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 Fig. 2, every choice task including a mix of positive and negative attributes will contain one or more dominant attributes. In the example in Fig. 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

[2] 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	–	$A_{4,1}^-$ $A_{4,2}^-$ $A_{4,3}^-$

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

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

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.

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.

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:

$$P\left(best = A_{k,l}\right) = P\left(V_{k,l} + \varepsilon_{k,l} > V_{m,n} + \varepsilon_{m,n}\right) \tag{1}$$

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):

$$P\left(best = A_{k,l_{k,s}}\right) = \frac{exp\left(V_{k,l_{k,s}}\right)}{\sum_{m=1}^{M} \exp\left(V_{m,n_{m,s}}\right)} \tag{2}$$

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:

$$V\left(A_{k,l}^{+}\right) + \varepsilon_{k,l}^{+} > V\left(A_{m,n}^{-}\right) + \varepsilon_{m,n}^{-} \tag{3}$$

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 Fig. 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:

$$P\left(best = A_{1,1}^{+}\right) = \frac{\exp\left(V\left(A_{1,1}^{+}\right)\right)}{\exp\left(V\left(A_{1,1}^{+}\right)\right) + \exp\left(V\left(A_{2,1}^{-}\right)\right) + exp\left(V\left(A_{3,2}^{-}\right)\right) + exp\left(V\left(A_{4,1}^{-}\right)\right)} \tag{4}$$

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.

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.

4. Simulation study

Section 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.

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 and 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 1 000 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/).

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.

The histograms in Fig. 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.

Dominance even affects the relative ranking of the positive attribute levels. Based on the utility values from Table 2 the correct

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

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**

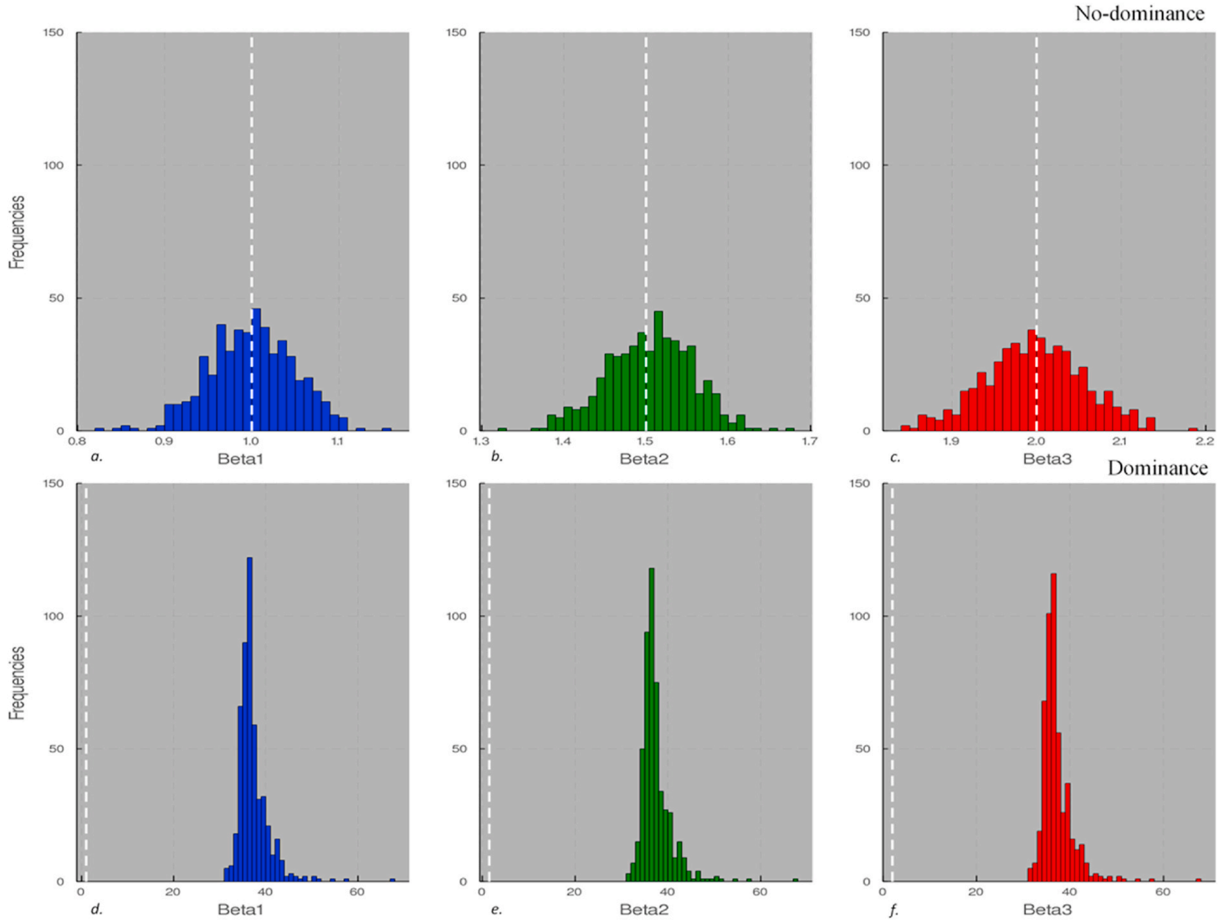

Fig. 3. Distribution of estimated utility values for attribute 1 V(A_1^+) and its levels in the simulations no-dominance (a–c) and dominance (d–f). Dashed white lines indicate the true utility values.

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.

To study what happens when there are multiple positive and negative attributes, we designed a second simulation study. In this

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%)

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 ($A^{+}_{2,3}$). Fig. 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.

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.[3] 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" 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 <u>not</u> being cured in Fig. 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 a number of 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

[3] 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.

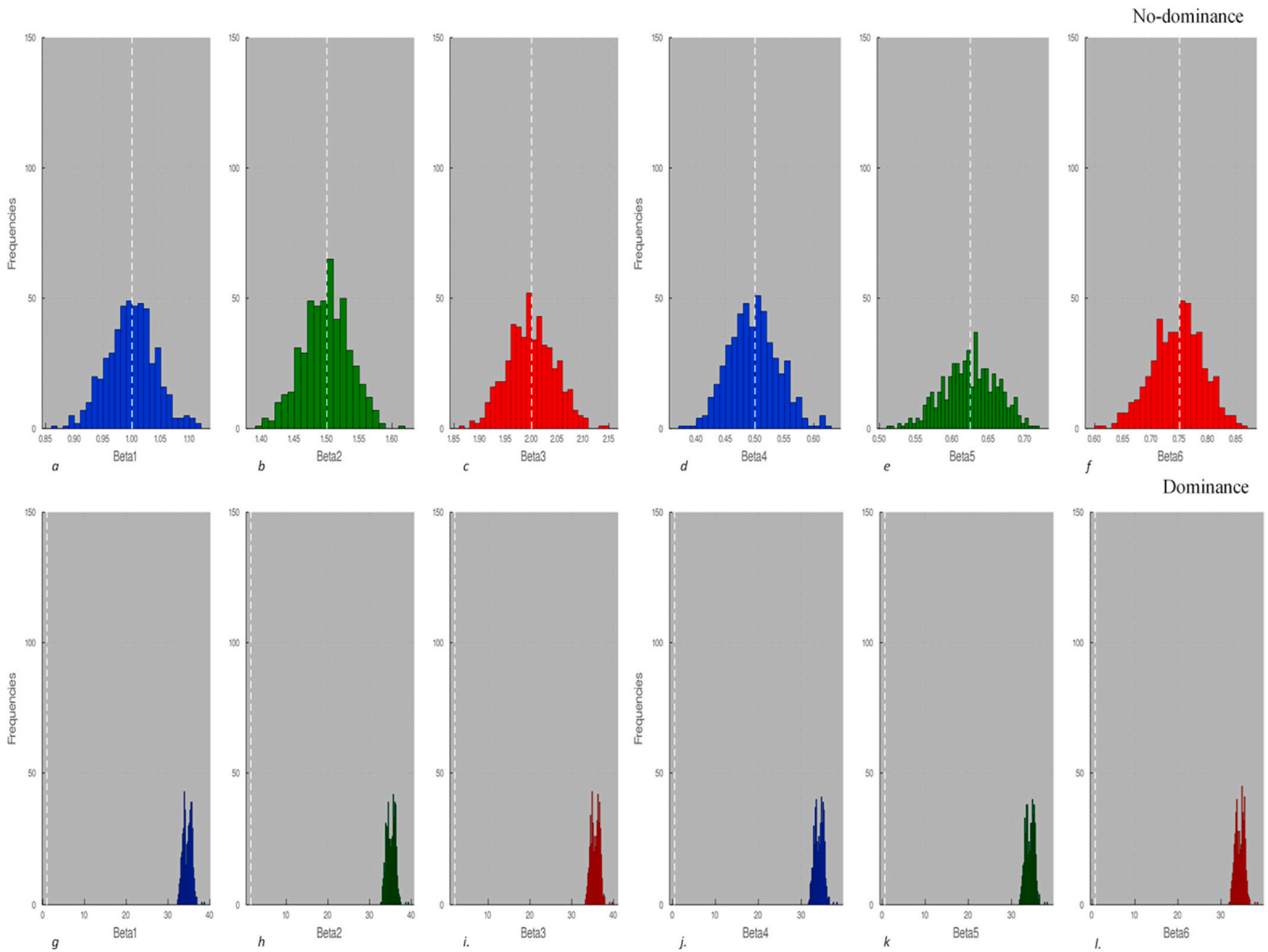

Fig. 4. Distribution of estimated utility values for the two positive attributes 1 V(A_1^+) and 2 V(A_2^+) in the no-dominance (a–f) and dominance scenarios (g–l). Dashed white lines indicate the true utility values.

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.

Financial support

Financial support for this study was provided both by the Research Excellence Initiative-Erasmus Choice Modelling Centre grant from the Erasmus University Rotterdam as well as the Patient Preferences in Benefit-Risk Assessments during the Drug Life Cycle (PREFER) project from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115966. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Disclaimer

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript. This text and its contents reflect the PREFER project's view and not the view of IMI, the European Union or EFPIA.

Author statement

V Soekhai: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration.

B Donkers: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision.

B Levitan: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision.

EW de Bekker-Grob: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

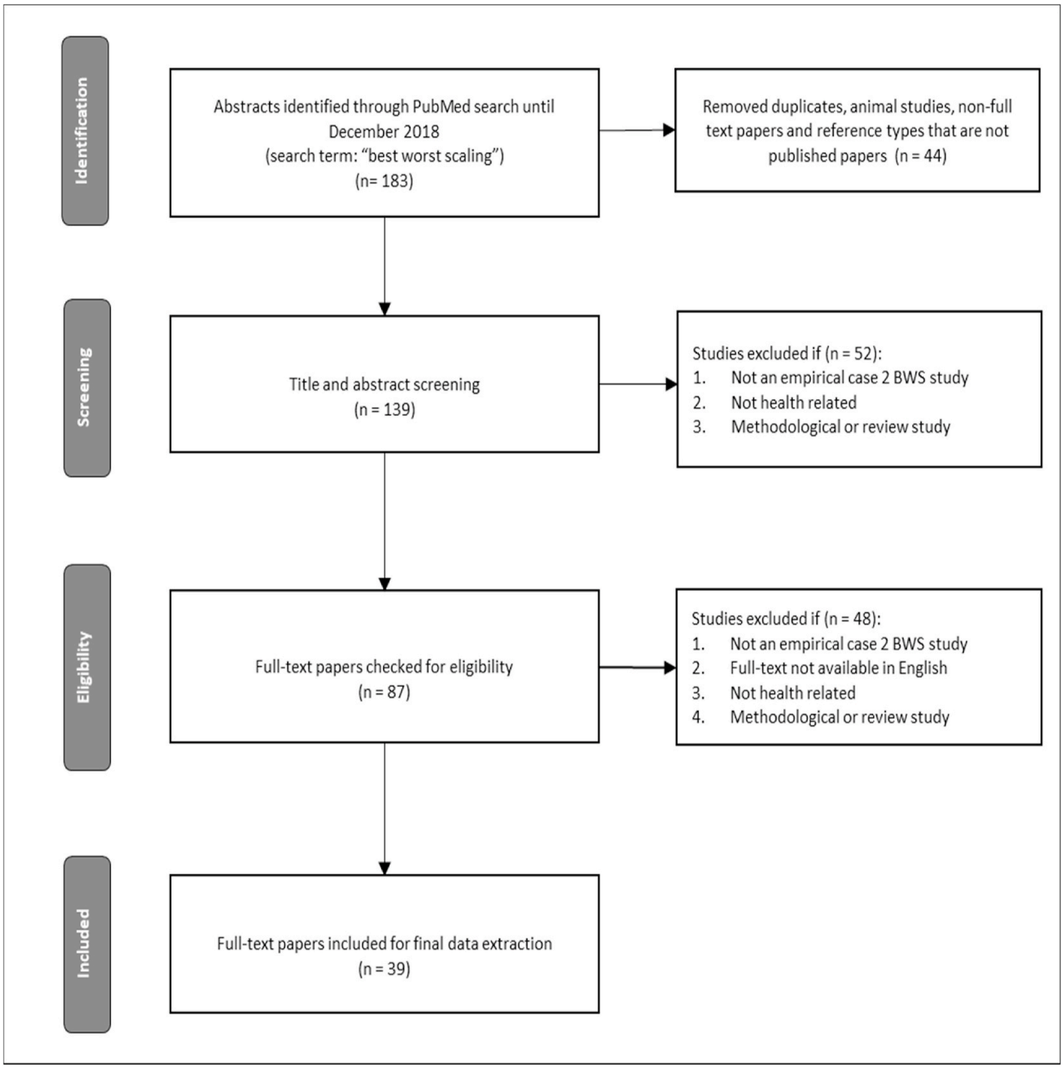

Fig. A. Flow diagram of scoping review to identify case 2 BWS studies.

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), 6 416–6 423. 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 and 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), 2129–2 136. 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), 2 765–2 773. 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), 1 655–1 661. 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.

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. Pharmacoecon. 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 Metrics 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. Ch. 8 in 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. Statistician 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. In: 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. In: 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., 2011. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc Sci Med. England 72, 1717–1727.

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. https://doi.org/10.1016/j.jpeds.2018.05.035 e1.

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 Health 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. In: Discrete Choice Methods with Simulation, second ed. 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 Health 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. Making 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. Simulat. Comput. 50 (7), 1989–1997. https://doi.org/10.1080/03610918.2019.1589511.