International Journal of Health Preference Research 2018, 1:3-16
http://dx.doi.org/10.21965/IJHPR.2018.001

Open Access Full Text Article

RESEARCH ARTICLE

Comparing different experimental designs for best-worst scaling choice experiments: the case of asthma control

[Comparaison de différents modèles expérimentaux pour les exercices de choix avec échelle best-worst: le cas du contrôle de l'asthme]

Mehdi Najafzadeh[1]
Wendy J. Ungar[2,3]
Anahita Hadioonzadeh[4]
Nicole Tsao[5]
Larry D. Lynd[5,6]

[1] Division of Pharmaco-epidemiology & Pharmaco-economics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School
[2] Program of Child Health Evaluative Sciences, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning
[3] The Institute for Health Policy, Management & Evaluation, University of Toronto
[4] Department of Health Policy, Management, and Behavior, State University of New York at Albany
[5] Collaborations for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences, University of British Columbia
[6] Centre for Health Evaluation and Outcome Sciences, Providence Health Research Institute

Correspondence:
Wendy J. Ungar, Program of Child Health Evaluative Sciences, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, 11th floor, 686 Bay Street, Toronto, ON, M5G 0A4, Canada

Email:
wendy.ungar@sickkids.ca

Article received:
27 March 2018

First response:
17 April 2018

Article accepted:
10 May 2018

Abstract: Best-worst scaling (BWS) experiments are useful for assessing preferences for health treatments. The objective was to compare the effects of an orthogonal main effects plan (OMEP) experimental design to a balanced incomplete block design [BIBD]) on preferences for pediatric asthma control. Five attributes were included in separate OMEP and BIBD questionnaires: Night-time symptoms, Wheezing or tightening of chest, Changing medication, Emergency visits, and Participation in physical activities. Convenience samples of parents with a child with asthma and adolescents with asthma recruited in Toronto, Canada, were randomly assigned to BIBD or OMEP. Preference weights were compared using conditional logit regression. Spearman's rank order test was used to assess agreement between the order of preference weights between the two designs. A total of 96 and 101 respondents completed the OMEP and BIBD questionnaires, respectively. Substantial agreement was observed between the order of preference weights in the two designs (ρ_Spearman=0.925; p-value<0.0001). Some differences were observed in the magnitude of preference weights with BIBD coefficients demonstrating a wider range within attributes compared to OMEP. In both designs, no Night time symptoms was the most preferred and ten Emergency room visits the least preferred attribute level. Preferences for different levels of Changing Medication had the smallest variation in both designs. The results suggest that using BIBD instead of OMEP for designing Case 2 BWS experiments may result in small but significant differences in preference estimates.

Keywords: best-worst scaling, discrete choice experiment, health preference, experimental design, asthma control, child.

Résumé : Les exercices de choix avec échelle Best-Worst (BWS) sont utiles pour évaluer les préférences de soins de santé. L'objectif était de comparer les résultats d'un plan expérimental à effets principaux orthogonaux (PEPO) à celui d'un plan expérimental en blocs équilibrés incomplets (PBEI) sur l'étude des préférences pour le contrôle de l'asthme à l'âge pédiatrique. Cinq caractéristiques ont été incluses dans les questionnaires distincts des approches PEPO et PBEI : Les symptômes nocturnes, la respiration sifflante ou gêne respiratoire, le changement de médication, les visites à l'urgence et la participation à des activités physiques. Un échantillon de convenance composé de parents ayant un enfant asthmatique et d'adolescents asthmatiques a été recruté à Toronto au Canada. Les participants ont été assignés aléatoirement dans l'une ou l'autre des approches PBEI ou PEPO. Les poids associés aux préférences ont été comparés en utilisant une régression logistique conditionnelle. Le test de classement de rang de Spearman a été utilisé pour évaluer le degré de corrélation dans le classement des préférences entre les deux approches. Au total, 96 répondants ont complété le questionnaire PEPO et 101 le questionnaire PBEI. Une corrélation substantielle a été observée au niveau du classement des poids associés aux préférences entre les deux approches (ρ_Spearman=0,925; p-value<0,0001). Quelques différences ont été notées dans l'amplitude des poids associés aux préférences avec des coefficients présentant un intervalle plus large avec l'approche PBEI comparativement à l'approche PEPO. Pour les deux approches, l'absence de symptômes nocturnes était la caractéristique préférée alors que dix visites à l'urgence était la moins appréciée.

La préférence pour différents niveaux de changement de médication avait la plus petite variation dans les deux approches. Les résultats suggèrent que l'utilisation de l'approche PBEI plutôt que PEPO pour la conception des méthodes BWS de type 2 peut conduire à de légères différences dans l'estimation des préférences.

Mots clés : échelle Best-Worst, choix expérimentaux discrets, préférence en santé, méthodologie expérimentale, contrôle de l'asthme, enfant.

Introduction

Best-Worst Scaling (BWS) is a type of discrete choice experiment (DCE) that is increasingly being used to elicit preferences for alternative options related to health states or health interventions [1]. In its simplest form, a BWS experiment is a systematic approach of gathering choice data to determine the relative position of a set of v items in the latent utility scale. In a BWS experiment, we present a series of choice tasks each containing a small subset (k) of v items (where k<v) and ask respondents to choose the best and the worst items among k possible items in each choice task. This process allows us to gather necessary and sufficient information about respondents' relative preferences for those v items. Similar to traditional DCE, BWS is based on random utility theory [1-6] and the underlying statistical assumption is that for a given set of k items, the probability of choosing a particular pair (as the best and worst options) is proportional to the distance between those two items in the latent utility scale [7]. The items can be a set of different "things" (referred to as Case 1), different attribute-levels of a "thing" (referred to as Case 2), or similar "things" with different profiles (referred to as Case 3) [8, 9]. An example of each of these three types of BWS in determining preferences for pediatric asthma control is presented in the Appendix.

Marley and Louviere [10] explored the mathematical properties of BWS experiments under Case 1. A balanced incomplete block design (BIBD) is a standard method to find partial factorial designs that ensure unbiased estimation of main effects and second order interactions for this type of BWS questionnaire [11]. Given v items, a BIBD can generate b blocks (i.e., choice tasks) each containing k items where: 1) each item appears at most once in each block; 2) each item appears in exactly r blocks; and 3) each possible pair of items appear in exactly λ blocks [11]. In BWS questionnaires, a "block" refers to a choice task, in contrast to a conventional discrete choice experiment where the term "block" refers to a subset of several choice tasks arrayed in a single questionnaire. Intuitively, the above conditions ensure that respondents are exposed to a balanced combination of different items as they answer the choice tasks in a BWS experiment. A BIBD enables us to design b choice tasks where only k items are presented in each choice task. Given that k can be much smaller than v, choosing the best and worst options among these k items in a choice task is cognitively easier and therefore, the choice data are prone to less random error.

In many practical situations however, items explored in a BWS experiment are comprised of levels of an attribute, such as a disease or treatment characteristic, and therefore, are classified as Case 2 BWS experiments. Unlike Case 1 where the items are independent, items in Case 2 are clustered within attributes. Marley et al. [12] describe Case 2 BWS experiments as those in which each choice task consists of a subset of attribute-levels. Given that a BIBD cannot account for this clustering effect, using a BIBD for designing Case 2 BWS experiments will result in having situations where some of the attributes appear more than once in a choice task (with different levels) while some other attributes may be absent in those choice tasks.

An orthogonal main effect plan (OMEP) is an alternative method for designing Case 2 BWS experiments. An OMEP imposes additional restrictions on the design to

account for clustering of items (i.e., attribute-levels) in groups (i.e., attributes). Therefore, using OMEP instead of BIBD ensures that each attribute appears exactly once in each choice task. Formally, given v items and g groups, an OMEP design can generate b blocks (i.e., choice tasks) each containing k items where: 1) each group has exactly one member in each block; and 2) each pair of items appear with proportional frequencies [11]. This can result in a more efficient instrument and reduce cognitive load on the respondent.

In general, however, it's easier to find an "off the shelf" BIBD design that is perfectly balanced and orthogonal. Finding an OMEP design that meets both of those criteria is often more difficult, due to the additional restrictions regarding clustering of items imposed by an OMEP design. Being able to utilize either type of design can increase the options available for choosing an appropriate experimental design. However, whether using BIBD for designing BWS experiments in Case 2 results in preference weight estimates that are similar to OMEP design has not been empirically explored. Therefore, understanding whether OMEP and BIBD designs can be used interchangeably in the design of Case 2 BWS experiments and whether they result in similar preference weight estimates is of great practical importance. In this study, we have used preference elicitation for pediatric asthma control using BWS as an empirical example to answer these questions. The study objective was to compare preference estimates from respondents randomly allocated to a BIBD or OMEP questionnaire and examine the impact of experimental design on the preference estimates in a Case 2 BWS experiment.

Methods

The study was approved by the Research Ethics Boards of The Hospital for Sick Children, Toronto, Canada, the University of British Columbia-Providence Health Care Research Institute, Vancouver, British Columbia, and the William Osler Health System, Brampton, Ontario, Canada (adolescent study only). All participants provided informed consent.

Sample selection and data collection

Convenience samples of parents with a young child with asthma and adolescents with asthma completed the BWS experiment from June 2011 to June 2012. Details of the recruitment have been previously published [13]. In brief, parent and adolescent respondents were recruited from an urban hospital outpatient asthma clinic, a community hospital asthma outpatient asthma clinic (adolescent sample only), a community hospital asthma education clinic (adolescent sample only), and a community-based asthma patient advocacy organization all located in the Greater Toronto Area, Ontario, Canada. Parents were eligible if they had a child between 2 and 12 years of age with a clinical diagnosis of asthma for whom maintaining asthma control was a present or past health issue, and they must have received at least one prescription for an asthma controller medication in the last year. The inclusion criteria were the same for the adolescent study, except participants had to be between 12 to 16 years old. Only one family could participate in either survey. Eligible patients were mailed a study package containing an information sheet, a randomly assigned BIBD or OMEP BWS questionnaire, and a parent-completed demographics and health questionnaire. The package also contained instructions for optional on-line questionnaire completion on our secure web site. Those opting to complete the questionnaire on the web were randomly assigned to complete a BIBD or OMEP questionnaire. Initiation of questionnaire completion indicated consent. Individuals who partially completed questionnaires were telephoned to collect missing data. Sawtooth software (Sawtooth Software, Sequim, WA) was used to design the questionnaires, to publish the questionnaire on the web, and to record responses into a secure database.

Of 336 study packages mailed to eligible candidates, a total of 206 were returned or completed on line (response rate 61.3%). The study packages included a BWS questionnaire followed by health and demographic questionnaire components and not all respondents completed all components. Complete health and demographic data were available for 104 OMEP and 99 BIBD respondents and complete BWS data were received from 96 OMEP and 101 BIBD respondents.

Experimental design

The five attributes and their corresponding levels related to the example of pediatric asthma control used in the design of the questionnaires are presented in Table 1. Using these five attributes, two separate questionnaires were developed based on OMEP and BIBD designs (see supplementary file – sample questionnaire). Sample choice tasks for OMEP and BIBD questionnaires are presented in Figure 1.

Table 1: Attributes and levels in the experimental comparison

Attribute	Level	Abbreviation
Night-time symptoms (NTS)	None	NTS0
	3 days per week	NTS3
	5 days per week	NTS5
Wheezing or tightening of chest (WTC)	No chest tightening or wheezing	WTC0
	Chest tightening or wheezing but it is manageable (does not worsen)	WTC1
	Chest tightening or wheezing and is bothersome (may worsen)	WTC2
Changing medication (CM)	No change needed	CM0
	More doses or adding on another asthma medication needed	CM1
	Adding oral steroids for 5 days needed	CM2
Emergency visits (EV)	No Emergency room visits	EV0
	1 Emergency room visit per year*	EV1
	4 Emergency room visits per year	EV4
	10 Emergency room visits per year	EV10
Participation in physical activities (PPA)	No physical activity limitations	PPA0
	2 limitations per month	PPA2
	10 limitations per month	PPA10

*Only used in the BIBD design.

The first BWS questionnaire was based on a symmetric OMEP design where five attributes (g=5) each with three possible levels were used to construct the questionnaire (therefore v=15 attribute-levels). The OMEP design was implemented in two versions. Each version consisted of nine (b=9) choice tasks where five attribute-levels (one level per attribute) were presented in each choice task (k=5). Table 2 shows the overall two-way frequencies of attribute-levels in the OMEP questionnaire including the number of times that each attribute-level was presented in the questionnaire (r=6) and the number of times that each possible pair of attribute-levels was presented in the questionnaire (λ). The cells with zero frequency in OMEP design indicate that each attribute was present exactly once in a choice task.

*Considering the following choices of attributes and their levels, please indicate which one you consider as the **most preferred (best)** and which one you consider as the **least preferred (worst)** attribute in asthma control.*

Please choose only <u>one</u> best and only <u>one</u> worst.

A. Orthogonal main effects plan

Best		Worst
☐	**Night-time symptoms:** *5 days per week*	☐
☐	**Wheezing or tightening of chest:** *Can be felt and is bothersome (may worsen)*	☐
☐	**Changing medication:** *To add oral steroids for 5 days*	☐
☐	**Emergency visits:** *None*	☐
☐	**Limitation of physical activities:** *2 limitations per month*	☐

B. Balanced incomplete block design

Best		Worst
☐	**Night-time symptoms:** *None*	☐
☐	**Night-time symptoms:** *5 days per week*	☐
☐	**Wheezing or tightening of chest:** *No chest tightening or wheezing*	☐
☐	**Changing medication:** *No changes to medication*	☐
☐	**Emergency visits:** *4 Emergency room visits per year*	☐
☐	**Emergency visits:** *10 Emergency room visits per year*	☐

Figure 1: Sample choice tasks

Table 2: Two-way frequencies of attribute levels in orthogonal main effects plan

	NS0	NS3	NS5	WT0	WT1	WT2	CM0	CM1	CM2	EV0	EV4	EV10	PA0	PA2	PA10
NS0	**6**	**0**	**0**	2	2	2	2	2	2	2	2	2	2	2	2
NS3	**0**	**6**	**0**	2	2	2	2	2	2	2	2	2	2	2	2
NS5	**0**	**0**	**6**	2	2	2	2	2	2	2	2	2	2	2	2
WT0	2	2	2	**6**	**0**	**0**	2	2	2	2	2	2	2	2	2
WT1	2	2	2	**0**	**6**	**0**	2	2	2	2	2	2	2	2	2
WT2	2	2	2	**0**	**0**	**6**	2	2	2	2	2	2	2	2	2
CM0	2	2	2	2	2	2	**6**	**0**	**0**	2	2	2	2	2	2
CM1	2	2	2	2	2	2	**0**	**6**	**0**	2	2	2	2	2	2
CM2	2	2	2	2	2	2	**0**	**0**	**6**	2	2	2	2	2	2
EV0	2	2	2	2	2	2	2	2	2	**6**	**0**	**0**	2	2	2
EV4	2	2	2	2	2	2	2	2	2	**0**	**6**	**0**	2	2	2
EV10	2	2	2	2	2	2	2	2	2	**0**	**0**	**6**	2	2	2
PA0	2	2	2	2	2	2	2	2	2	2	2	2	**6**	**0**	**0**
PA2	2	2	2	2	2	2	2	2	2	2	2	2	**0**	**6**	**0**
PA10	2	2	2	2	2	2	2	2	2	2	2	2	**0**	**0**	**6**

The matrix indicates the frequencies that each attribute-level appeared in the OMEP questionnaire (r=6) and the number of times that each possible pair of attribute-levels appeared together in the questionnaire. The cells with zero frequency indicate that each attribute was present exactly once in a choice task.

The second BWS questionnaire was designed using a BIBD. The closest feasible BIBD (a (16,6,2)) to our OMEP design required using 16 attribute-levels (v=16). Therefore, we used exactly the same five attributes and levels and added one additional level to one of the attributes. Attribute-level "Emergency visits: 1 emergency room visit per year" was added to the 15 OMEP attribute-levels to meet the required number of items (16 attribute-levels) for this particular BIBD. The BIBD then used 16 choice tasks (b=16), each presenting 6 attribute-levels (k=6). Table 3 shows the overall two-way frequencies of attribute-levels in the BIBD questionnaire including the number of times that each attribute-level was presented in the questionnaire (r=6) and the number of times that each possible pair of attribute-levels was presented in the questionnaire (λ=2). Unlike the OMEP design, the BIBD does not account for clustering of attribute-levels within attributes. Consequently, BIBD does not impose any restriction on co-appearance of different levels of an attribute in a choice task and therefore, an attribute could appear more than once or not at all in a choice task (Figure 1B).

Statistical analysis

Descriptive statistics of participants in the two study arms were estimated by analysis of demographic and child health data using SAS for Windows, version 9.2. Possible differences between the two arms were compared using t-tests for continuous variables and chi-square or Fisher's Exact test for categorical variables.

The choice data were coded and analysed separately for the OMEP and the BIBD questionnaires. Conditional logit regression was used to analyse the choice data using Latent Gold Choice, version 4.5. The coefficients for all attribute-levels were estimated relative to the last attribute-level (i.e., Limitation in physical activities: 10 limitations per month) which was chosen as the reference. The estimated coefficients of the conditional logistic model represent respondents' average preference weights for a given attribute-level relative to the reference attribute-level.

Log odds ratios for the two designs and their 95% confidence intervals were estimated from the regression coefficients. These ratios were used to account for scale differences between the two designs and

Table 3. Two-way frequencies of attribute-levels in balanced incomplete block design

	NS0	NS3	NS5	WT0	WT1	WT2	CM0	CM1	CM2	EV0	EV1	EV4	EV10	PA0	PA2	PA10
NS0	**6**	**2**	**2**	2	2	2	2	2	2	2	2	2	2	2	2	2
NS3	**2**	**6**	**2**	2	2	2	2	2	2	2	2	2	2	2	2	2
NS5	**2**	**2**	**6**	2	2	2	2	2	2	2	2	2	2	2	2	2
WT0	2	2	2	**6**	**2**	**2**	2	2	2	2	2	2	2	2	2	2
WT1	2	2	2	**2**	**6**	**2**	2	2	2	2	2	2	2	2	2	2
WT2	2	2	2	**2**	**2**	**6**	2	2	2	2	2	2	2	2	2	2
CM0	2	2	2	2	2	2	**6**	**2**	**2**	2	2	2	2	2	2	2
CM1	2	2	2	2	2	2	**2**	**6**	**2**	2	2	2	2	2	2	2
CM2	2	2	2	2	2	2	**2**	**2**	**6**	2	2	2	2	2	2	2
EV0	2	2	2	2	2	2	2	2	2	**6**	**2**	**2**	**2**	2	2	2
EV1	2	2	2	2	2	2	2	2	2	**2**	**6**	**2**	**2**	2	2	2
EV4	2	2	2	2	2	2	2	2	2	**2**	**2**	**6**	**2**	2	2	2
EV10	2	2	2	2	2	2	2	2	2	**2**	**2**	**2**	**6**	2	2	2
PA0	2	2	2	2	2	2	2	2	2	2	2	2	2	**6**	**2**	**2**
PA2	2	2	2	2	2	2	2	2	2	2	2	2	2	**2**	**6**	**2**
PA10	2	2	2	2	2	2	2	2	2	2	2	2	2	**2**	**2**	**6**

The matrix indicates the frequencies that each attribute-level appeared in the BIBD questionnaire (r=6) and the number of times that each possible pair of attribute-levels appeared together in the questionnaire (λ=2). Unlike the OMEP design, the BIBD does not account for clustering of attribute-levels within attributes. Consequently, in BIBD different levels of the same attribute could appear more than once or not at all in a choice task.

to enable valid comparisons [14]. The log odds ratios for all attribute levels were estimated relative to the last attribute level (i.e., PPA10). Therefore, a positive (negative) coefficient indicates a greater (lesser) preference compared to the reference attribute level.

The statistical model describing the relationship between underlying preferences and observed choices in a BWS experiment are described based on the following formulation. The utility of respondent *i* choosing a particular pair *j* (where b is best and w is the worst item) in the choice task *t can* be formulated as:

$$U_{jt}^{i}(b,w) = V_{jt}^{i} + \varepsilon_{jt}^{i}$$

Where V_{jt}^{i} is systematic and ε_{jt}^{i} is random component of utility. Systematic part of utility can be described as:

$$V_{jt}^{i} = \beta_{b}^{i} * I(b) - \beta_{w}^{i} * I(w)$$

Where *I(.)* is an indicator function and is equal to 1 if the argument is true and 0 otherwise. Based on these formulations, probability of choosing a particular pair *j* in the choice task *t* can be formulated as:

$$P_{jt}^{i} = \frac{\exp(V_{jt}^{i})}{\sum_{l}\exp(V_{lt}^{i})}$$

Where l represents all possible best-worst pairs available in the choice task *t*.

The potential difference of the estimated preference weights between the two arms can be attributed to: 1) effect of BIBD versus OMEP design or 2) differences between respondents in the two arms. By randomizing the respondents to either of the two designs, we expect to minimize respondent-related differences in preference weights. Finally, we used Spearman's rank order test to assess agreement between the two designs in terms of order of preference weights.

Results

The demographic statistics in the two arms of the experiment are reported in Table 4. The sample represented children with moderate to severe asthma, based on reported rates of night-time asthma

Table 4: Baseline characteristics of participants in the BWS experiment

Characteristic	OMEP (n=104)		BIBD (n=99)		p-value
	n	%	n	%	
Child's age parent sample (years), mean (SD)	7.6 (2.5)		6.9 (2.9)		0.1942
Child's age teen sample (years), mean (SD)	13.5 (1.2)		14.1 (1.1)		0.0262
Male child	68	65.4	56	56.6	0.1977
Parents born in Canada	58	55.8	54	54.6	0.8609
Parental education					0.6847
University or college degree/diploma	69	66.4	66	66.7	
Some university or college	15	14.4	20	20.2	
Completed high school or less	20	19.2	13	13.1	
Family has a drug benefits plan	83	79.8	73	73.7	0.5845
Annual household income					0.4370
Less than $10,000 to $59,999	30	28.8	25	25.3	
$60,000 to $120,000	33	31.7	21	21.2	
Greater than $120,000	21	20.2	27	27.3	
Not sure or prefer not to respond	13	26.0	13	25.5	
Asthma attacks in last 6 mo, mean (SD)	2.2 (3.7)		3.0 (7.2)		0.2826
History of other respiratory conditions					
Pneumonia	27	26.0	30	30.3	0.4914
Bronchitis	20	19.2	29	29.3	0.0940
Croup	20	19.2	22	22.2	0.5989
Child catches cold or respiratory infections more often than other children	58	55.8	65	65.7	0.1496
Symptom frequency in the last month					0.8620
None	29	27.9	23	23.2	
1-2 times a month	29	27.9	30	30.3	
One to three times per week	29	27.9	31	31.3	
One to four times per day	14	13.5	14	14.1	
Other	3	2.9	1	1.0	
Night-time asthma symptoms in last month	57	54.8	51	51.5	0.0821
≥ 1 family doctor visit in last 6 months	29	27.9	29	29.3	0.9419
≥ 1 pediatrician visit in last 6 months	22	21.2	21	21.2	0.9095
≥ 1 specialist visit in last 6 months	71	68.3	67	67.7	0.4042
≥ 1 emergency room visit in last year	29	27.9	33	33.3	0.8529
≥ 1 hospital admission in last year	10	9.6	15	15.2	0.7793
Received asthma management or action plan	75	72.1	70	70.7	0.8209
Asthma medications used in last year					0.9532
BD + ICS or BD + AL	58	55.8	58	59.2	
BD + ICS + AL	28	26.9	22	22.5	
Oral steroid	6	5.8	7	7.1	
Other	12	11.5	11	11.2	
Missing value	2	4.0	0	0.0	

Abbreviations: SD = standard deviation; mo = month; BD = bronchodilator; ICS = inhaled corticosteroid; AL = anti-leukotriene.

symptoms in the previous month, emergency room visits in the previous year, and hospital admissions for asthma in the previous year. Use of an asthma controller medication was reported in the majority of respondents. None of the health and demographic variables were statistically different between the two groups suggesting successful randomization.

The estimated log odds ratios for the two designs and their 95% confidence intervals are reported in Table 5.

Table 5: Estimated log odds ratios in OMEP and BIBD designs

Attribute level		OMEP			BIBD			Mean Difference	p-value
		Mean	95%CI		Mean	95%CI			
Night time symptoms									
None	NTS0	3.51	3.18	3.85	3.73	3.47	3.99	-0.22	0.32
3 days/ week	NTS3	1.31	0.99	1.62	0.85	0.61	1.09	0.46	0.02*
5 days/ week	NTS5	0.52	0.20	0.83	-0.35	-0.58	-0.13	0.87	<0.01*
Wheezing/ chest tightening									
None	WTC0	3.19	2.85	3.53	3.49	3.23	3.74	-0.30	0.17
Manageable	WTC1	2.14	1.81	2.47	1.64	1.37	1.92	0.49	0.02*
Bothersome	WTC2	0.53	0.22	0.83	-0.18	-0.41	0.05	0.70	<0.01*
Changing medication									
None	CM0	2.64	2.30	2.97	2.55	2.29	2.82	0.08	0.70
More doses/add medication	CM1	1.57	1.24	1.90	0.66	0.40	0.91	0.92	<0.01*
Add oral steroids for 5 days	CM2	1.42	1.08	1.76	0.37	0.13	0.61	1.05	<0.01*
Emergency room visits									
None	EV0	3.36	3.03	3.70	3.19	2.93	3.45	0.17	0.43
1 per year	EV1	-	-	-	1.17	0.93	1.42	-	-
4 per year	EV4	0.38	0.07	0.68	-0.03	-0.27	0.21	0.41	0.04*
10 per year	EV10	-0.48	-0.78	-0.17	-1.00	-1.22	-0.78	0.53	0.01*
Participation in physical activities									
No limitation	PPA0	3.22	2.88	3.56	3.21	2.95	3.46	0.01	0.96
2 per month	PPA2	1.31	0.98	1.65	1.20	0.93	1.46	0.12	0.60
10 per month	PPA10	0.00	-	-	0.00	-	-	-	-

Log-odds ratios with upper and lower 95% Confidence Internals are presented for each attribute level. P-values are for t-tests with df=95 and assuming a normal distribution of estimated preference weights ($t = \frac{\beta_1 - \beta_2}{\sqrt{se_1^2 + se_2^2}}$; $\approx \min(n_1, n_2) - 1$). * = P < 0.05

The directions and signs of all estimated preference weights in each of the designs were in concordance with our prior expectations with regard to ordering of preference weights for levels within each attribute (Figure 2). For example, the log odds ratio of NTS0 ("Night time symptoms: None") was larger than the log odds ratio for NTS5 ("Night time symptoms: 5 days per week"), showing a rational direction of preferences. The same pattern was observed for all attributes.

Overall, there were similarities between log odds ratios in the two designs. In both designs, the log odds ratio for NST0 ("Night time symptoms: None") had the largest positive coefficient and the log odds ratio for EV10 ("Emergency room visits: 10") had

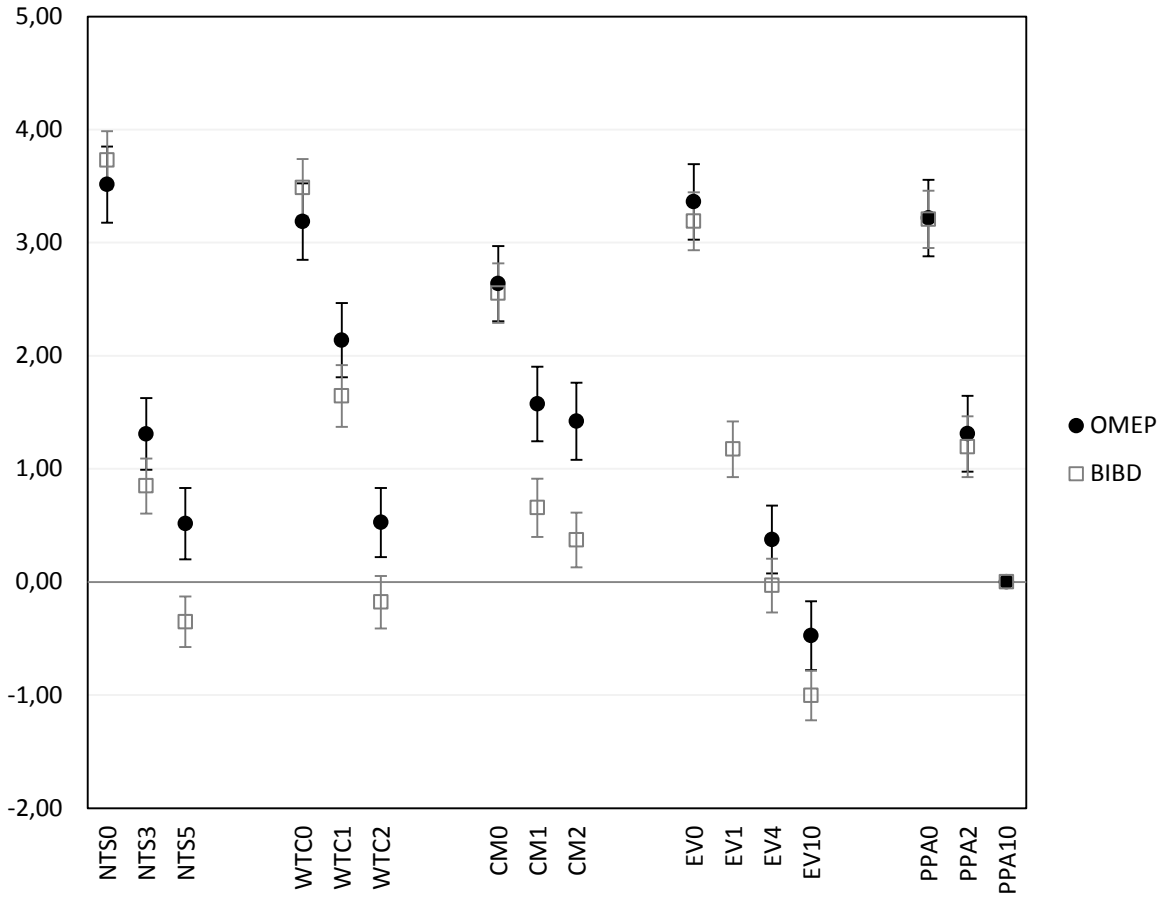

Figure 2: Estimated log odds ratios in OMEP and BIBD designs. The preference weights for each level within each attribute are plotted. Circles indicate the estimated preference weights for each attribute-level using the OMEP questionnaire. Squares indicate the estimated preference weights for each attribute-level using the BIBD questionnaire.

the largest negative coefficient. In addition, in both designs log odds ratios for different levels of CM ("Changing Medication") had the smallest range of variation (i.e., smallest positive and negative coefficients). Spearman's rank order test suggested a large and statistically significant agreement between the order of log odds ratios in the two designs (ρ_Spearman=0.925; p-value<0.0001). But this agreement was less than unity due to rank reversals in some of preferences, in particular in the estimated log odds ratio for WTC0, CM1, CM2, WTC2, and NTS5 in the BIBD (Figure 3).

Despite major similarities, some differences in the magnitude of log odds ratios between the two designs were observed. The differences between log odds ratios for each attribute level and the corresponding p-values based on a t-test are presented in Table 5. In general, the estimated log odd ratios in the BIBD were slightly "stretched" in both positive and negative directions. For example, the log odds ratio for NTS3 was 0.46 smaller in the BIBD compared with the OMEP design (diff=0.46; p-value=0.02) and the coefficient for NTS5 was 0.87 smaller in the BIBD compared with corresponding values in the OMEP design (diff= 0.87; p-value<0.01). This effect also resulted in significant differences between coefficients of WTC1 (diff=0.49; p-value=0.02), WTC2 (diff=0.70; p-value<0.01), CM1 (diff=0.92; p-value<0.01), and CM2 (diff=1.05; p-value<0.01), EV4 (diff= 0.41; p-value=0.04), and EV10 (diff= 0.53; p-value=0.01) in the two designs. None of the differences between other coefficients in the two designs was statistically significant.

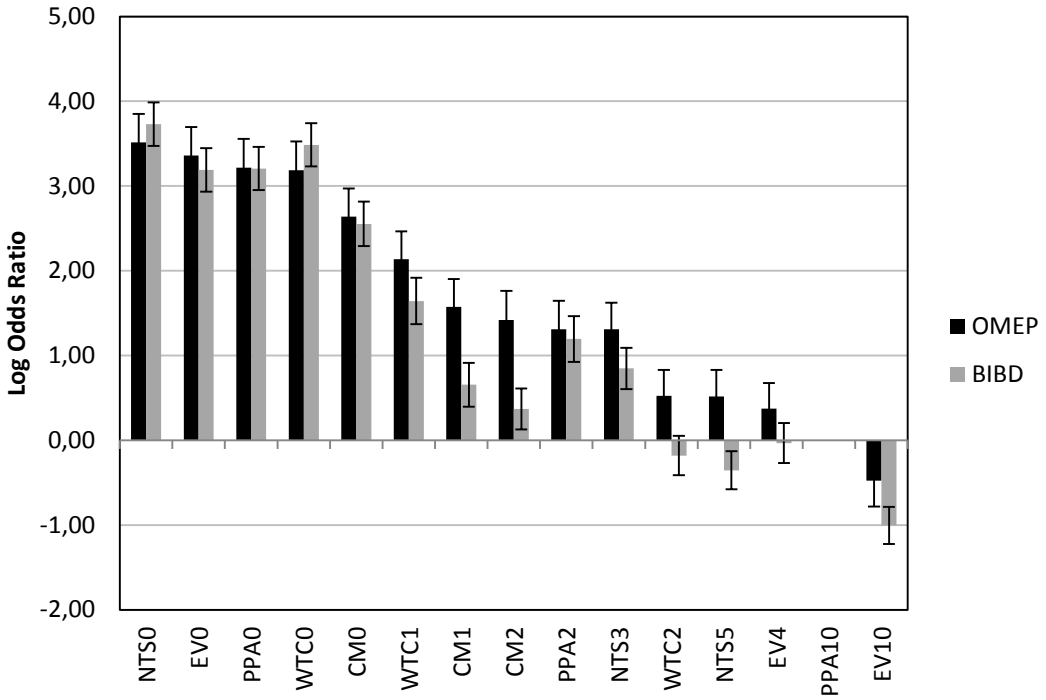

Figure 3: Order of log odds in OMEP versus BIBD designs. The preference weights for each attribute-level are plotted in descending order. The black bars indicate estimated preference weights for each attribute-level using the OMEP questionnaire. The grey bars indicate the estimated preference weights for each attribute-level using the BIBD questionnaire. Spearman's rank order test suggested a large and statistically significant agreement between the order of preference weights in the two designs ($\rho_{Spearman}=0.925$; p-value<0.0001).

Discussion

Using preferences of parents and adolescents for asthma control as a case study, we investigated the effect of statistical design (i.e., OMEP vs. BIBD) on the estimated preference weights elicited using a Case 2 BWS experiment. Our results suggested that the log odds of the preference weights were comparable in terms of both magnitude and rank order in two designs. However, these agreements were not perfect and existence of meaningful differences between the designs rejected the null hypothesis of achieving no difference in preference estimates using either design. Therefore, using BIBD design for designing Case 2 BWS should be done with caution as it appeared to impact the results in our example. In the absence of a gold standard, the choice between the two designs can be made based on other considerations such as improving response efficiency by administering fewer choice tasks and fewer options within each choice task.

Our findings suggest that using OMEP in designing Case 2 BWS experiments may be considered as the preferred approach. An OMEP design uses full profiles in each choice task (i.e., all attributes are present in a choice task) and unlike a BIBD, restricts the appearance of more than one level per attribute in a choice task. These properties, in particular when the attributes have ordered levels, can eliminate obvious choices from the set of choice tasks and therefore increases efficiency in the data collection process. For example, a choice task in a BIBD that includes both "night-time symptoms: none" and "night-time symptoms: 3 days per week" renders it as a less efficient method for gathering preference information. In our case study, the OMEP BWS questionnaire had only 9 choice tasks (compared with 16 choice tasks in the BIBD BWS questionnaire). In addition,

OMEP respondents were presented with only 5 items in each choice task compared to 6 items in the BIBD. Therefore, respondents' cognitive burden can be reduced using the OMEP design while the confidence intervals of coefficients in the two designs were comparable given the similar sample sizes. Overall, our results suggest that using OMEP is a more efficient approach to designing the BWS questionnaires in Case 2 BWS experiments in health care. Nevertheless, we expect that the differences between OMEP and BIBD preference estimates may become negligible when the attributes have no clearly ordered levels.

To the best of our knowledge, this is the first study that compared the effect of BWS questionnaire design on preference estimates. By using a randomized experiment, we were able to parse out the effect of design on the preference estimates. Traditional DCE is the method of choice when the aim is to quantify marginal rates of substitution between different attributes to estimate metrics such as maximum acceptable risk or willingness to pay. However, as a newer approach for preference elicitation, BWS has some methodological advantages over traditional DCE when the goal is to determine a relative ranking of attributes. First, BWS is a more efficient way to collect information about individuals' preferences about health as they state both the most and least preferred options in a given choice task. Second, indicating the items that are at the extreme of the preference scale (i.e., selection of the most preferred option and least preferred option) is a relatively easier task for respondents and results in choices with better consistency and smaller random error. Third, in contrast to traditional DCE, BWS models allow estimation of preference weights for all but one item relative to the last item [7, 10]. Thus preferences for items can be compared across disparate attributes. One important advantage of Case 2 BWS over standard DCE is its ability to measure relative ranking of all attribute-levels compared to levels within a single attribute. Based on Spearman's rank order test, we observed a large and statistically significant agreement between the order of preference weights for attribute-levels in the two designs. Therefore, both OMEP and BIBD designs performed equally well in achieving this goal.

Increasing interest in BWS technique warrants further attention to the methodological aspects of questionnaire design and data analysis [15-17]. In considering BWS, one must recognize that all of the models for analysis of BWS data (e.g., paired model, marginal model, or conditional logit) [7] rely heavily on respondents' choice frequency of best-worst pairs. Consequently, using designs that expose respondents to an unfair presentation of candidate pairs can result in biased and inefficient estimation of preference weights. As such, using an appropriate questionnaire design is a crucial step in developing a BWS choice experiment and can directly affect the accuracy and precision of the preference estimates. Although there are a large number of OMEP and BIBD designs, only a handful of those designs are practical for use in BWS questionnaire design in health care. In reality, finding an OMEP design or BIBD that can match a pre-defined number of attributes and levels is often challenging and sometimes impossible. We believe that obtaining unbiased estimates should be the primary goal in a BWS experiment design. Achieving this goal necessitates having flexibility in selection of the number of attributes and levels so that they can achieve perfect balance and orthogonality.

Conclusions

In our case study, the two BWS experiments designed using OMEP and BIBD produced comparable preference weight estimates in a sample of parents of children with asthma and adolescents with asthma. However, small but significant differences in preferences between the two designs suggest that OMEP can be marginally advantageous compared to BIBD in the design of Case 2 BWS questionnaires where

both attributes and attribute levels are provided in the choice task. Given the direct effect of BWS questionnaire design on bias and precision of preference estimates, the number of attributes and levels during the questionnaire design should be selected based on the feasibility of an OMEP or BIBD to address the health question.

Acknowledgments

The clinical collaboration of Dr. Sharon Dell and the technical assistance of Ms. Salma Lalji are gratefully acknowledged.

Funding

Financial support for this study was provided by a grant from the AllerGen Network of Centres of Excellence and by in-kind support from The Hospital for Sick Children Research Institute, the Asthma Society of Canada, and the Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences, University of British Columbia. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Conflicts of interest

The authors declare no potential or perceived conflicts of interest.

References

[1] 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.

[2] Louviere JJ, Hensher DA, Swait JD. Stated Choice methods: analysis and application. Cambridge: Cambridge University Press; 2000.

[3] Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing discrete choice experiments: Do optimal designs come at a price? J Consum Res 2008;35(2):360-75.

[4] Luce RD. The choice axiom after twenty years. J Math Psychol 1977;15(3):215-33.

[5] Marley AAJ. Random utility models and their applications: Recent developments. Math Soc Sci 2002;43(3):289-302.

[6] McFadden D. Econometric models for probabilistic choice among products. J Bus 1980;53(3):S13-S29.

[7] Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. J Health Econ 2007;26(1):171-89.

[8] Lancsar E, Louviere J, Donaldson C, Currie G, Burgess L. Best worst discrete choice experiments in health: methods and an application. Soc Sci Med 2013;76(1):74-82.

[9] Marley A, Pihlens D. Models of best-worst choice and ranking among multiattribute options (profiles). J Math Psychol 2012;56(1):24-34.

[10] Marley A, Louviere J. Some probabilistic models of best, worst, and best-worst choices. J Math Psychol 2005;49(6):464-80.

[11] Green PE. On the design of choice experiments involving multifactor alternatives. Journal of Consumer Research 1974;1(2):61-8.

[12] Marley AAJ, Flynn TN, Louviere JJ. Probabilistic models of set-dependent and attribute-level best-worst choice. Journal of Mathematical Psychology 2008;52:281-96.

[13] Ungar WJ, Hadioonzadeh A, Najafzadeh M, Tsao NW, Dell S, Lynd LD. Quantifying preferences for asthma control in parents and adolescents using best–worst scaling. Respir Med 2014;108:842-51.

[14] Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinational logit models. Journal of Marketing Research 1993;30(3):305.

[15] Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: An empirical comparison using social care data. Soc Sci Med 2011;72(10):1717-27.

[16] Severin F, Schmidtke J, Muhlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: A pilot study comparing best-worst scaling and discrete-choice experiments. Eur J Hum Genet 2013;21(11):1202-8.

[17] Louviere JJ, Flynn TN, Marley AAJ. Best-Worst Scaling. Theory, Methods and Applications. Cambridge, UK: Cambridge University Press; 2015.

Appendix. Examples of choice tasks in Case 1, Case 2, and Case 3 BWS experiments.

Case 1. A sample choice task in a Case 1 BWS experiment

Best		Worst
☐	**Night-time symptoms**	☐
☐	**Wheezing or tightening of chest**	☐
☐	**Changing medication**	☐
☐	**Limitation of physical activities**	☐

Case 2. A sample choice task in a Case 2 BWS experiment

Best		Worst
☐	**Night-time symptoms:** 3 days per week	☐
☐	**Wheezing or tightening of chest:** No chest tightening or wheezing	☐
☐	**Changing medication:** To add oral steroids for 5 days	☐
☐	**Emergency visits:** 10 Emergency room visits per year	☐
☐	**Limitation of physical activities:** 2 limitations per month	☐

Case 3. A sample choice task in a Case 3 BWS experiment

	Scenario A	**Scenario B**	**Scenario C**	**Scenario D**
Night-time symptoms	3 days per week	5 days per week	None	3 days per week
Wheezing or tightening of chest	No chest tightening or wheezing	None	Is bothersome	None
Changing medication	To add oral steroids for 5 days	Adding oral steroids for 3 days needed	No change needed	Adding oral steroids for 5 days needed
Emergency visits	10 Emergency room visits per year	0 Emergency room visits per year	1 Emergency room visits per year	4 Emergency room visits per year
Limitation of physical activities	2 limitations per month	None	10 limitations per month	2 limitations per month
Night-time symptoms	3 days per week	5 days per week	None	3 days per week
Please indicate which scenario is the BEST in your opinion.	☐	☐	☐	☐
Please indicate which scenario is the WORST in your opinion.	☐	☐	☐	☐