Social Science & Medicine 76 (2013) 74–82

Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

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

Best worst discrete choice experiments in health: Methods and an application

Emily Lancsar [a,*], Jordan Louviere [b], Cam Donaldson [c], Gillian Currie [d], Leonie Burgess [e]

[a] Centre for Health Economics, Monash University, Clayton, VIC 3800, Australia
[b] Centre for the Study of Choice, University of Technology, Sydney, Australia
[c] Centre for Social Business and Health, Glasgow Caledonian University, United Kingdom
[d] Departments of Paediatrics and Community Health Sciences, University of Calgary, Canada
[e] School of Mathematical Sciences, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Available online 2 November 2012

Keywords:
Canada
Best worst discrete choice experiments
Best worst scaling
Ranking
Sequential best worst
Heterogeneity
Cardiac arrest

ABSTRACT

A key objective of discrete choice experiments is to obtain sufficient quantity of high quality choice data to estimate choice models to be used to explore various policy/clinically relevant issues. This paper focuses on a relatively new form of choice experiment, 'Best Worst Discrete Choice Experiments' (BWDCEs) and their relevance to health research as a new way to meet such an objective. We explain what BWDCEs are, how and when to apply them and we present several analytical approaches to model the resulting data. We demonstrate this preference elicitation approach in an empirical application exploring preferences of 898 members of the general public in Edmonton and Calgary, Canada for treatment of cardiac arrest occurring in a public place and show the gains achieved compared to traditional analysis of first best data. We suggest that BWDCEs are a valuable way to investigate preferences in the health sector and discuss implications for task design, analysis and areas for future research.

Introduction

The broad research, clinical and policy applicability of discrete choice experiments (DCEs) has been embraced in the health sector. DCEs have been used with patients, providers, policy makers and the general public to explore preferences and values in relation to health and health care, and more recently in health outcome measurement and economic evaluation (De Bekker, Ryan, & Gerard, 2012; Lancsar & Louviere, 2008a). DCEs involve asking respondents to choose the best of several alternatives offered in choice sets, thereby obtaining one choice observation per set. A key objective of DCEs is to obtain sufficient quantity of high quality choice data to estimate choice models that can be used to explore various policy/clinically relevant issues. Historically there have been two main ways to obtain more choice observations in DCEs: 1) increase the sample size; and/or 2) ask sampled respondents to evaluate more choice sets, choosing their most preferred alternative in each. The former clearly has cost implications while the latter can impose task burdens on respondents. More recently a third way was proposed (Louviere et al., 2008): 3) instead of more choice sets, ask more questions per set. Best worst DCEs (BWDCEs) are designed to do just that by asking respondents to choose not only the best option in each choice set, but also the worst option, followed by the best from the remaining options and so on until an implied preference ordering of the options is obtained.

BWDCEs are highly relevant to health related research and are likely to prove useful where DCEs are currently used since they provide richer information on relative preferences between alternatives and significantly larger amounts of choice data which typically results in gains in statistical efficiency as we later show. BWDCEs also can help achieve a given target number of choice observations while reducing sample sizes and should prove particularly useful in applications with small sample sizes, arising from budget constraints or small populations from which to sample. Examples of the latter abound in health settings such as samples from specialist physicians, policy makers, patients with rare diseases, etc. In such cases, BWDCE is likely to open up new research areas allowing one to study stated preferences that may not be feasible with data generated from a standard DCE. Finally, when combined with an appropriate experimental design, the extra choice information can be used to estimate models for single individuals, providing a new way to capture heterogeneity (Lancsar & Louviere, 2008b; Louviere et al., 2008).

BWDCEs are one of three types of best worst scaling (BWS) (Finn & Louviere, 1992), and differ from the type of BWS most commonly used in the health sector to date, which ask respondents to choose

* Corresponding author.
E-mail address: Emily.Lancsar@monash.edu (E. Lancsar).

0277-9536/$ – see front matter
http://dx.doi.org/10.1016/j.socscimed.2012.10.007

the best and worst attribute levels in a single profile/option (see Flynn, Louviere, Peters, & Coast, 2007). While BWDCEs are gaining interest, there are relatively few papers in health or more generally and no comprehensive guide. BWDCE were introduced by Louviere et al. (2004) and first used in health by Lancsar and Louviere (2005) and Lancsar, Louviere, and Flynn (2007), the latter reports the results of models based only on the first (best) data (that is, by ignoring subsequent worst and best data collected in each choice set). Subsequently Fiebig, Knox, Viney, Haas, and Street (2011) also employed BWDCE data collection, but only used the first best data. Brown, Pashos, Joshi, and Lee (2011) recently used a BWDCE task but provide little information on their design or methods of analysis. Such choice tasks also have been used in the broader literature, including by Louviere et al. (2008), to estimate individual level models for pizza and juice products, and Scarpa, Notaro, Louviere, and Rafelli (2011), in an application to tourism in alpine grazing commons.

The growing interest but paucity of literature in this area provides the opportunity to make a useful contribution by focussing attention on the relevance of the BWDCE method in health research, and by providing guidance on undertaking and analysing a BWDCE. We present a menu of analytical approaches including a model (the sequential best worst MNL) that accounts for the best worst data generation process. The latter represents another contribution since in analysing BWDCE data to date, researchers have largely ignored the best worst structure of the data. We also show how preference and variance heterogeneity can be accounted for in the analysis of best worst data. We apply and demonstrate BWDCE in a health application exploring preferences for treatment of cardiac arrest occurring in a public place (outside a hospital and/or home) in Calgary and Edmonton, Alberta, Canada. In addition to exploring preferences in this area, another objective of the empirical analysis is to demonstrate analytical methods that can be used to estimate models from BWDCE data.

The next section provides background to BWDCEs as a preference elicitation method and their relation to BWS tasks more generally. We discuss several ways to analyse such data that account for varying amounts and composition of preference information in the third section. The fourth section describes the empirical study exploring preferences for treatment for cardiac arrest in a public place, with results presented in the fifth section. We conclude with a discussion of the methods, results and areas for further research.

BWDCEs and their relation to best worst scaling

BWDCEs are a form of BWS (Finn & Louviere, 1992), of which there are three types: best worst object scaling; best worst attribute scaling; and BWDCEs, referred to as BWS Cases 1, 2 and 3, respectively (Flynn, 2010). Each involves asking respondents to choose the best and worst (most and least preferred) from a set of three or more items.

Best worst object scaling (Case 1) (Finn & Louviere, 1992) involves asking respondents to choose the best and worst object, but unlike traditional DCEs, the objects are not decomposed into effects associated with attributes. It has been used, for example, by Finn and Louviere (1992) to investigate opinions about food supply policy and their relation to other public policy issues and by Auger, Devinney, and Louviere (2007) to investigate ethical beliefs about product related issues such as recycled packaging and social factors such as human rights. Case 1 recently was used in health to measure the principles that citizens thought were most and least important to guide health care reform in Australia (Louviere & Flynn, 2010). Proofs of the mathematical and measurement properties of Case 1 are in Marley and Louviere (2005).

Best worst attribute scaling (Case 2) (Louviere, Finn, & Timmermans, 1994) involves evaluation of single profiles (attribute level combinations) one-at-a-time. Each profile represents a choice set from which respondents choose the best and worst attribute *levels*. Thus, choices are made within, not between alternatives. The mathematical and measurement properties of Case 2 are in Marley, Louviere, and Flynn (2008). Case 2 allows researchers to measure attribute levels on a common scale, overcoming the inability to compare attribute impacts directly in traditional DCEs (Flynn et al., 2007; Lancsar et al., 2007; McIntosh & Louviere, 2002). Work with vulnerable populations suggests Case 2 tasks may be easier for some respondents than traditional DCEs (Flynn et al., 2007). However, a potential disadvantage of Case 2 is it does not allow typical policy analysis associated with traditional DCEs, such as predicted probability analysis or estimation of welfare measures because such analysis requires discrete choices between rather than within alternatives.

Case 2 is the most frequently used type of BWS in health to date. For example, by Szeinbach, Barnes, McGhan, Murawiski, and Corey (1999) to elicit utility weights for EQ-5D health states and by McIntosh and Louviere (2002) to measure preferences for dental treatment options, amongst others. Subsequently, Case 2 was discussed in more detail in a user guide by Flynn et al. (2007) with several recent applications (Coast et al., 2008; Flynn, Louviere, Peters, & Coast, 2008; Lancsar et al., 2007; Potoglou et al., 2011; Ratcliffe et al., 2011).

BWDCE (Case 3) is the type of BWS closest to traditional DCEs. Like standard DCEs, BWDCEs involve respondents making repeated choices between alternatives offered in choice sets, each described by a number of attributes. BWDCEs are designed to elicit extra preference information per choice set by asking respondents not only to choose the best option but also to sequentially choose the worst option, potentially followed by choice of best of the remaining options and so on until an implied preference ordering is obtained over all alternatives in a set. For a choice set containing J alternatives respondents can be asked to make up to J-1 sequential best and worst choices. For example, if a choice set contains 5 alternatives (A, B, C, D, E), asking for best, worst, next best and next worst per set provides a complete implied preference order.

Compared to a standard DCE, respondents are asked more choice questions in each choice set but as they already have evaluated all alternatives in a set to choose the best, the burden of asking for worst and so on is less than asking them to evaluate additional choice sets and asking for the best in each. For example, in the empirical study discussed below respondents considered 16 choices sets and chose best, worst, second best and second worst from the 5 alternatives presented per set, producing 64 observations per respondent. To obtain the same number of observations from a standard DCE would require each respondent considering 64 choice sets and selecting the best in each. Similarly, BWDCEs offer advantages in terms of cognitive effort and consistency compared with traditional ranking tasks (Louviere et al., 2008; Marley & Louviere, 2005; Scarpa et al., 2011). The cognitive advantages lie in making the task of providing a ranking over options easier, eliciting the preference order in stages as a reiterated set of best worst choices (Scarpa et al., 2011). Marley and Louviere (2005) note that best worst tasks seem to be easy for people to complete and suggest that such tasks take advantage of a person's propensity to identify and respond more consistently to extreme options.

There are relatively few applications of Case 3 tasks in health. The mathematical properties of BWDCEs are starting to be considered (Marley, 2011), and the empirical application in this paper shows that experimental design theory for traditional DCEs also can be used with BWDCEs. More generally, designs for specific types of best worst tasks have begun to be considered (Vermeulen,