Too many attributes: A test of the validity of combining discrete-choice and best–worst scaling data

Jing Zhang [a,1], F. Reed Johnson [b,1,*], Ateesha F. Mohamed [c,1], A. Brett Hauber [d]

[a] Department of Bioresource Policy, Business and Economics, University of Saskatchewan, Canada
[b] Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University, 2400 Pratt Street, Durham, NC 27705, USA
[c] GHEOR, Specialty Medicine, Bayer Healthcare Pharmaceuticals Inc., Whippany, NJ, USA
[d] RTI Health Solutions, RTI International, Research Triangle Park, NC 27709, USA

ARTICLE INFO

ABSTRACT

Background: Best-practice guidelines for stated-preference methods suggest there is a limit to the number of attributes respondents can reliably evaluate. This study explores a cost-effective solution to combining elicitation formats from a single study to obtain more preference information from a given sample while limiting respondents' cognitive burden.
Methods: A stated-preference survey combined both discrete-choice experiment (DCE) and best–worst scaling (BWS) elicitation formats to Alzheimer's disease caregivers. DCE questions elicited attribute-level preferences for one subset of attributes, and object-case BWS elicited overall relative attribute importance for another subset of attributes, with two overlapping attributes in both designs. Two alternative joint models combined preferences from the BWS and DCE data. One model controlled for confounding between response-error variance and preference parameters in the DCE model, and the other did not.
Results: About 400 caregivers completed the survey. We estimated attribute-level preference parameters for 17 attributes, 9 of which were directly estimated using the DCE data, and 8 of which were extrapolated based on the overall relative importance estimated using the object-case BWS data. Results from both joint models and individual models indicate that relative preferences from the two question formats were the same up to a scale factor.
Conclusion: Our results suggest that combining DCE and object-case BWS is a cost-effective solution to the need for more information when study resources are limited. Moreover, for these data at least, researchers' concerns about serious confounding between DCE model estimates and response-error variance appear unwarranted.

© 2015 Published by Elsevier Ltd.

1. Background

Drug-approval decisions by regulatory agencies require evidence on treatments' clinical benefits and risks. There is an increased interest in quantifying patients' risk tolerance for given clinical benefit and patient-reported outcome (PRO)

---

* Corresponding author.
  E-mail addresses: jzhang26@gmail.com (J. Zhang), reed.johnson@duke.edu (F. Reed Johnson), ateesha.mohamed@bayer.com (A.F. Mohamed), abhauber@rti.org (A.B. Hauber).
  [1] These individuals were employed by RTI International at the time this study was conducted.

benefits (European Medicines Agency, 2010; Center for Devices and Radiological Health, 2012). Stated-preference studies, particularly discrete-choice experiments (DCE), or choice-format conjoint analysis, have increasingly been used to provide evidence on patient preferences in health outcomes research (Center for Devices and Radiological Health, 2012; Johnson et al., 1998; Hauber et al., 2010; Bridges et al., 2007; Scarpa et al., 2009). Best-practice guidelines for DCEs caution against eliciting preference information for a large number of attributes (Marshall et al., 2010; Bridges et al., 2011). However, clinical trials that incorporate PRO outcomes can have a large number of endpoints, including multiple measures of efficacy, safety, side effects, toxicities, and multidimensional PRO measures such as health-related quality of life and functional abilities in activities of daily living.[2]

While DCEs have been widely used to quantify patient preferences for treatment outcomes and benefit–risk tradeoffs in health and other applications in the last decade (Marshall et al., 2010; Bridges et al., 2011), the popularity of the best–worst scaling (BWS) method has increased in the last few years to elicit health preferences (Marshall et al., 2010; Marley and Louviere, 2005; Marley et al., 2008; Flynn et al., 2008; Louviere and Flynn, 2010; Flynn et al., 2008; Flynn, 2010). The rising popularity of BWS is related to the purported ease with which respondents can identify the best and the worst from a list of items, the flexibility of asking different types of BWS questions to elicit different types of preference information, and the relative ease of designing BWS instruments and analyzing the data using simple methods such as net counts (Louviere and Flynn, 2010; Flynn et al., 2008; Flynn, 2010; Flynn et al., 2007). The simplest form of the BWS question is object-case BWS where items are attributes rather than attribute levels or full profiles. The object-case BWS format asks respondents to rate the best and the worst attributes in a list of items. Because of the relative ease of answering BWS questions, the number of attributes that can be included in a BWS question are less restrictive than DCE questions, especially when the overall relative importance of attributes are of research interest, not within-attribute level tradeoffs. For example, Louviere and Flynn (Louviere and Flynn, 2010) used the object-case BWS to elicit preferences for 15 healthcare reform principles from a sample of 204 respondents in Australia.

However, an obvious disadvantage of this approach is that, unlike DCE formats, object-case BWS does not provide preference information for each attribute level, but only for attributes as a whole. For example, for 3 attributes with 3 levels each we can estimate 6 dummy-coded DCE parameters (two attribute-level parameters for each attribute, normalizing the omitted category at zero). Object-level BWS data support estimating only 2 dummy-coded parameters for the overall attribute importance.

The primary objective of the study was to quantify caregiver benefit–risk tradeoffs among attributes that measure treatment efficacy and safety outcomes, while a secondary objective was to quantify relative preferences for activity items in the Disability Assessment for Dementia (DAD) PRO instrument. The DAD was developed to assess disabilities in individuals with cognitive deficits. It has 10 different dimensions, such as hygiene, continence, and eating; each dimension is measured by several different DAD activity items, many of which are highly correlated.[3]

Given sample-size constraints, the cognitive limitations of respondents' abilities to accurately evaluate a large number of attributes, this study implemented a strategy to obtain more information by using two different preference-elicitation formats to elicit preferences for 17 attributes in a single, cognitively tractable stated-preference survey. We elicited preferences for the efficacy and safety attributes using a DCE elicitation format, and preferences for the DAD activity attributes using an object-case BWS format. To evaluate the effect of elicitation format on estimated preferences, 2 of the 10 DAD activity items also were included in the DCE design. We then tested the hypothesis that data from the two elicitation formats are consistent with the same relative preferences except for a constant scale difference. Failure to reject this hypothesis would allow researchers to employ a similar data-enrichment strategy to obtain preference estimates for longer attribute lists than usually is feasible in DCE studies.

2. Methods

2.1. Survey sample, question formats, and experimental designs

A stated-preference web survey elicited AD caregiver benefit–risk tradeoff preferences for 7 attributes that measure treatment efficacies and safety outcomes and for 10 DAD activity items.[4] Prior to survey administration, the survey was pretested in face-to-face interviews with 20 caregivers.[5] A total of 403 US caregivers were eligible and consented to participate in the study in the summer of 2011. The sample size is well above the average for DCE applications in health

---

[2] Over the period 1997–2002, of 214 new drugs approved by the United States Food and Drug Administration (FDA), about half mentioned multi-dimensional PRO instruments.

[3] The 10 dimensions are hygiene, dressing, continence, eating, meal preparation, telephoning, going on an outing, finance and correspondence medications and leisure and housework.

[4] Caregivers of patients with Alzheimer's disease (AD) were invited to participate in the study, because patients with advanced AD are cognitively impaired and cannot provide informed consent,

[5] The survey was administered to informal caregivers of AD patients in the United States (US) and Germany. Prior to survey administration, the survey was first pretested with 20 caregivers in the US and 5 caregivers in Berlin, Germany. Due to space limitations, only the results from the US data are presented in this paper.