VALUE IN HEALTH 18 (2015) 217–223

Available online at www.sciencedirect.com

ScienceDirect

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

Preference-Based Assessments

Learning and Satisficing: An Analysis of Sequence Effects in Health Valuation

Benjamin M. Craig, PhD[1,], Shannon K. Runge, MA[1], Kim Rand-Hendriksen, PhD[2], Juan Manuel Ramos-Goñi, BSc[3], Mark Oppe, PhD[4]*

[1]*Health Outcomes and Behavior, Moffitt Cancer Center and University of South Florida, Tampa, FL, USA;* [2]*Department of Health Management and Health Economics, University of Oslo, Oslo, Norway;* [3]*Canary Islands Health Service (SESCS), Santa Cruz de Tenerife, Spain;* [4]*Institute for Medical Technology Assessment, Erasmus University Rotterdam, Rotterdam, The Netherlands*

ABSTRACT

Objective: To estimate the effect of sequence on response precision and response behavior in health valuation studies. **Methods:** Time trade-off (TTO) and paired comparison responses from six health valuation studies—four US, one Spanish, and one Dutch—were examined (22,225 respondents) to test whether task sequence influences response precision (e.g., rounding), response changes, and median response times. Each study used a computer-based instrument that randomized task sequence among a national sample of adults, age 18 years or older, from the general population. **Results:** For both TTO and paired comparisons, median response times decreased with sequence (i.e., learning), but tended to flatten after the first three tasks. Although the paired comparison evidence demonstrated that sequence had no effect on response precision, the frequency of rounded TTO responses (to either 1-year or 5-year units) increased with sequence. **Conclusions:** Based on these results, randomizing or reducing the number of paired comparison tasks does not appear to influence response precision; however, generalizability, practicality, and precautionary considerations remain. Overall, participants learned to respond efficiently within the first three tasks and did not resort to satisficing, but may have rounded their TTO responses.
Keywords: health valuation, paradata, preferences, QALY, response precision, sequence effects, time trade-off.

Introduction

Most economic evaluations summarize effectiveness using preference weights on a quality-adjusted life-year (QALY) scale, as recommended by numerous health technology assessment agencies. Such QALY weights may be from societal or patient perspectives and derived using a wealth of preference elicitation tasks (e.g., best-worst scaling). Although valuation research has a well-established history, the use of online computer-based surveys for health valuation offers an array of new capabilities, such as quota-sampling at the task level; paradata on respondent behavior, device, and browser; and other interactive technologies. Compared with interview, postal, or telephone surveys, online computer-based experiments increase control in the randomization of tasks, while reducing cognitive burden and minimizing missing data and other data collection errors and biases.

Although online instruments typically randomize the order of presentation of tasks, response precision and behavior may change with sequence. For example, when a respondent is shown two alternatives and asked, "Which do you prefer?" he or she may take longer or change his or her responses on initial pairs while becoming acquainted with the valuation task as compared with later pairs. Furthermore, a respondent's attention may wane in later pairs, leading to satisficing (i.e., expediting selection among alternatives to minimize effort), reducing response precision [1,2]. This article examines whether response precision and response behavior vary with the number of tasks completed (i.e., sequence effect) in health valuation studies for two types of valuation tasks, time trade-off (TTO) and paired comparisons.

Understanding the relationship between response precision and task sequence guides the number of tasks to be included in a valuation study, informs weights that place a greater emphasis on earlier or later tasks, and justifies the randomization of task sequence. Although studies have attempted to identify respondents who randomize all responses (i.e., shufflers and satisficers) [3], few studies to date have examined the effect of sequence on response precision in health valuation [4].

Sequence effects have been identified in other forms of discrete choice experiments (DCEs) as a type of ordering effect specifically related to the order in which choice sets are presented (i.e., position-dependent order effects) [5]. This type of order effect differs from those related to the order or position of

* *Address correspondence to:* Benjamin M. Craig, Moffitt Cancer Center, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612.
E-mail: benjamin.craig@moffitt.org.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2014.11.005

attributes within a choice set [5–7]. Experimental design, such as the layout of questions, the number of attributes, and the number of tasks, can influence ordering effects and response time [8–10]. A key example in survey research is the primacy effect or the tendency for respondents to choose the first reasonable answer to a survey question (e.g., first response option in a list of potential answers) [6,11]. This weak form of satisficing leads to nonrandom response; expedites response with minimum effort; reduces response quality and time; and is commonly cited by experimenters to justify randomization and reduction in the number of attributes, scenarios, and tasks [12].

A wealth of studies have examined order effects in terms of perception and salience [5,7,9,10,13–17], although the results have been somewhat inconsistent. For example, some evidence suggests that the order of attributes affects choice [5,7], yet other studies did not find this effect [9,14,18]. In addition, the number and complexity of task sets within an experiment may induce order effects through respondent fatigue or boredom [19]. Evaluating the association between participant response behaviors (i.e., response times and changes) and task sequence has the potential to provide valuable insight regarding the influence of study design.

In complement to evidence on response precision, we examine response behaviors (i.e., response times and changes) that may indicate learning and added deliberative effort beyond that which is needed to satisfy the task requirements. Typically, response behavior is examined at the questionnaire level (e.g., the amount of time it takes a respondent to complete all tasks). In addition to evaluating response behavior at the questionnaire level, computerized software offers a unique opportunity to examine response behaviors at the level of individual questions (e.g., the amount of time it takes to complete a single task set or a series of different task sets). A better understanding of response behavior at each of these levels can aid in the interpretation of the empirical association between sequence and response precision and in the improvement of survey design (e.g., cognitive burden).

The present study contributes to an innovative evaluation of client-side paradata. Client-side paradata is the information recorded in Web surveys by the respondent's computer (e.g., the number of times and locations of mouse clicks on a computer screen). Unlike server-side paradata, which refers to data management processes, client-side information allows researchers to interpret participant response behaviors in terms of changed responses (CRs) and response time at the level of individual questions [20]. Evaluating response behavior patterns at such a specific level contributes to our knowledge of how sequence influences preferences. In this secondary analysis of health valuation data, we examine sequence effects, specifically whether response precision and response behavior vary with the number of tasks completed.

Methods

Preference Elicitation

In a paired comparison, respondents are asked, "Which do you prefer?" given two health episodes, and their choices define the relative value between these episodes. An original TTO task is more involved, using an adaptive series of paired comparisons based on either time with no health problems or "immediate death." Specifically, each TTO begins with a paired comparison in which the respondent must first decide whether the health episode is preferred to immediate death. If so, an adaptive series of paired comparisons is presented to determine the number of years with no health problems that is equivalent to the health episode (i.e., better-than-dead indifference statement). If the respondent prefers immediate death, an alternative series of paired comparisons is completed to identify a worse-than-dead indifferent statement. The original adaptation procedure [21–23] is like a dose-response study in that it increases the duration of problems within an episode until it is equivalent to immediate death (e.g., how much poison is needed until it kills you). Thus, the TTO exercise is a matching task that produces an equivalence statement regardless of whether the original paired comparison response is better or worse than death.

Data

To test the effect of sequence on response precision and behavior, we examined paired comparisons and TTO responses from six health valuation studies—four US, one Spanish, and one Dutch—totaling 259,318 responses from 22,225 respondents who completed 17 to 37 tasks [2,24–27]. Table 1 summarizes the characteristics of these six studies. All studies used a computerized instrument that randomized task sequence using national samples of adults from the general population. For the US-based studies, respondents completed a set of paired comparisons trading improvements in health-related quality of life (HRQOL)

Table 1 – Health valuation studies*.

Study title	Dates	No.	First set of tasks	Second set of tasks
Patient Reported Outcomes Measurement Information System (PROMIS) Valuation Study - United States [2]	March–July 2012	7557	6 lifespan pairs	24 health pairs
EQ-5D-5L Valuation Study - Spanish	May–July 2012	986	10 time trade-offs	7 health state pairs†
Child Health Valuation Study - US, Wave 1 [24]	July–August 2012	2008	6 lifespan pairs	31 health pairs
EQ-5D-5L Valuation Study - Dutch [27]	September–October 2012	1052	10 time trade-offs	7 health state pairs†
Child Health Valuation Study - United States, Wave 2 [24]	January–February 2013	2147	12 lifespan pairs	18 health pairs
Women's Health Valuation Study - United States [25]	April 2013	3397	8 lifespan pairs	22 health pairs
Measurement and Valuation of Health Study - United States [26]	November–December 2013	5078	8 lifespan pairs	22 health pairs

EQ-5D-5L, five-level EuroQol five-dimensional questionnaire.
* Each wave of the US Child Health Valuation Study is shown separately because of changes in the valuation tasks.
† Unlike health and lifespan pairs, health state pairs do not describe duration in the health state.