International Journal of Technology Assessment in Health Care, **18:1** (2002), 139–144.

RESEARCH NOTE

INVESTIGATING THE STRUCTURAL RELIABILITY OF A DISCRETE CHOICE EXPERIMENT WITHIN HEALTH TECHNOLOGY ASSESSMENT

Julie Ratcliffe

RTI—Health Solutions

Louise Longworth

Health Economics Research Group and Brunel University

Abstract

This paper investigates the structural reliability of a discrete choice experiment within health technology assessment. Two versions of a discrete choice experiment, in the form of a self-completion questionnaire, were randomly administered to two samples of women who had recently given birth as part of an exercise to determine women's preferences for alternative modes of intrapartum care. In the first questionnaire, two of the attributes had only their highest and lowest levels included, while in the second questionnaire all three levels for these two attributes were included. The levels included for all other attributes remained the same throughout both questionnaires. The evidence relating to the structural reliability of the discrete choice experiment in this context was mixed. The results indicated that the relative importance of the two attributes in which the levels were varied increased as the number of levels for these attributes increased. However, the relative importance of the attributes in which the levels were not varied remained relatively stable throughout. The results provide evidence in support of a psychological effect whereby respondents place more importance upon specific attributes as the number of levels for these attributes increases. It is recommended that further research of both a qualitative and quantitative nature should be undertaken to assess the potential importance (or otherwise) of a psychological effect relating to the number of levels chosen for attributes within discrete choice experiments in health technology assessment.

Keywords: Structural reliability, Health care, Discrete choice experiments, Conjoint analysis

Discrete choice experiments (DCEs) represent one form of conjoint analysis—a stated preference technique for establishing patient and public preferences—which is gaining popularity in technology assessments in health care, as evidenced by the increasing number of DCEs being undertaken and reported upon in recent years within the health care sector (2;5;6;8;9;10). An important methodologic issue of interest for DCEs in health care concerns the structural reliability of the measurement approach over attribute set. One way in which

This study was undertaken while Julie Ratcliffe was based at the Health Economics Research Group at Brunel University. The authors would like to thank members of the Health Economists' Study Group, in particular Stirling Bryan and Emma McIntosh, for their comments on an earlier version of this paper. The authors are particularly grateful to Professor Mary Boulton for her assistance in the development and administration of this study. The authors would also like to thank Stella Axom, Lorna Branczik, and Peter Edwards for their help in identifying the sample and questionnaire administration.

Table 1. Attributes and Levels Included in the Study

Attribute	Levels
CONT: Continuity of contact with the same medical staff[b]	None[a] Yes: limited Yes: high
LOCAT: Location of delivery	Hospital labor ward[a] Maternity unit with a home-like environment Home
PAINREL: Availability of pain relief	Gas and air only Gas and air plus birthing pool All types[a]
DECIS: Decision making[b]	By medical staff[a] Jointly by medical staff and woman By woman
TRANSFER: Probability of transfer during labor	None[a] Yes: low probability Yes: high probability

[a] Indicates levels of attributes for constant comparator.
[b] Indicates attributes where the number of levels was varied between exercises.

this can be examined is to vary the number of levels for some of the attributes while keeping the levels for all remaining attributes constant. This issue was examined as part of a project that investigated women's preferences for alternative modes of maternity care during the intrapartum stage.

METHODS

Following the convening of two focus group meetings with women who had recently given birth in inner London, five main attributes were identified as potentially important in determining choices between alternative modes of intrapartum care. The chosen attributes and the levels assigned to each of these are shown in Table 1. A constant comparator was used throughout the exercise, which most closely resembled the characteristics associated with a hospital birth. The remaining attribute levels were formulated into scenarios using the computer software package SPEED version 2.1 (1). In order to assess structural reliability, two scenario formulation tasks were undertaken. In the first task (exercise A), the attributes CONT and DECIS had only their highest and lowest levels included, while in the second task (exercise B) all three levels for these two attributes were included. The levels included for all other attributes remained the same throughout both tasks. Each of the two versions of 16 scenarios chosen by the SPEED software was randomly split into two groups of eight scenarios, and four versions of the questionnaire were produced with eight pairwise comparisons (constant comparator versus alternative scenario) in each version.

Given the repeated measurement aspect of the data (whereby multiple responses are obtained from the same individual), the random effects probit model was chosen as an appropriate model by which to analyze the data (4). The four versions of the main questionnaire were then randomly allocated to all women who had received a home birth at two hospitals in the London area over the time period from May 1998 through April 1999 inclusive ($n = 192$) as well as a 50% larger sample of hospital births ($n = 290$).[1] One reminder was sent to nonrespondents after a time interval of approximately 4 weeks.

Structural reliability was assessed in three ways. First, for every respondent, tests were carried out to assess whether any of the attributes were dominant (7). Evidence in

https://doi.org/10.1017/S0266462302108130 Published online by Cambridge University Press

favor of structural reliability in this context would be indicated if the number of dominant respondents for each attribute were broadly similar across both exercises, since this would indicate that the incentive to trade between attributes would be similar for respondents to both exercises. Second, for nondominant respondents, a simple comparison of the random effects probit models for exercises A and B was undertaken. The functional forms for the two models are:

Model A

$$\Delta Ui_a = \propto_0 + \propto_1 C_a + \propto_2 L_a + \propto_3 P_a + \propto_4 D_a + \propto_5 D_a + \propto_6 T_a + e_{ia} + u_1$$

Model B

$$\Delta Ui_b = \propto_0 + \propto_1 C_b + \propto_2 L_b + \propto_3 P_b + \propto_4 D_b + \propto_5 D_b + \propto_6 T_b + e_{ib} + u_1$$

where ΔUi is the change in utility in moving from hospital to home-based care within each choice; $i = 1, \ldots, n$ is the number of nondominant respondents to the survey; a or $b = 1, \ldots, 8$ is the number of choices posed in each exercise; C, L, P, D, T are the difference in levels within each choice for the CONT, LOCAT, PAINREL, DECIS, and TRANSFER attributes, respectively; $\propto_0 - \propto_6$ are the model coefficients; e_i is the random measurement error due to differences among observations; and u_1 is the random measurement error due to differences among nondominant respondents. Both models were specified using a constant term in order to investigate the presence of a systematic tendency to choose the left or right option, which in this context may indicate an underlying preference for the location of intrapartum care.[2]

Third, a single model was specified using the dummy variable approach, whereby the data from exercises A and B were pooled and the following model (Model C) estimated:

Model C

$$\begin{aligned}\Delta Ui_a = {} & \propto_0 + \propto_1 C_a + \propto_2 L_a + \propto_3 P_a + \propto_4 D_a + \propto_5 D_a + \propto_6 T_a + \propto_7 [A_{ia} C_a] \\ & + \propto_8 [A_{ia} L_a] + \propto_9 [A_{ia} P_a] + \propto_{10} [A_{ia} D_a] + \propto_{11} [A_{ia} T_a] + e_{ia} + u_1\end{aligned}$$

where $A_{ia} = 1$ for observations from exercise A and 0 for observations from exercise and all other parameters are as defined previously. In this case, the difference in the models for the two exercises is indicated by the statistical significance of the differential constant ($\propto_0$) and the differential slope coefficients ($\propto_7 + \propto_8 + \propto_9 + \propto_{10} + \propto_{11}$). The null hypothesis of no difference is:

$$\propto_7 = 0; \quad \propto_8 = 0; \quad \propto_9 = 0; \quad \propto_{10} = 0; \quad \propto_{11} = 0.$$

RESULTS

A total of 257 usable questionnaires were returned, giving an overall response rate of 55%. Analysis of the response patterns of individuals to each exercise revealed three distinct groups. For exercise A, 36% (n = 44) of respondents exhibited dominant preferences for home births, 31% (n = 38) exhibited dominant preferences for hospital births, and the remaining 33% (n = 39) were prepared to trade between scenarios on the basis of the changing levels of the attributes presented. For exercise B, 35% (n = 48) of respondents exhibited dominant preferences for home births, 36% (n = 49) exhibited dominant preferences for hospital births, and the remaining 29% (n = 39) were prepared to trade between scenarios on the basis of the changing levels of the attributes presented. Table 2 summarizes the

Table 2. Descriptive Characteristics of Nondominant Respondents

Characteristic	Exercise 1 (n = 39)	Exercise 2 (n = 39)	*p* Value
Age	17–25 6 (15.4%) 26–35 25 (64.1%) >35 8 (20.5%)	17–25 8 (20.5%) 26–35 23 (59.0%) >35 8 (20.5%)	0.822
Highest education level obtained	No qual. 2 (5.1%) O level 7 (17.9%) A level 8 (20.5%) Degree 18 (46.2%) Postgrad 2 (5.1%) Other 2 (5.1%)	No qual. 2 (5.1%) O level 9 (23.1%) A level 7 (17.9%) Degree 17 (43.6%) Postgrad 3 (7.7%) Other 1 (2.6%)	0.530
Where was last baby delivered?	Home 16 (41.0%) Hospital 20 (51.3%) Other 3 (7.7%)	Home 18 (46.2%) Hospital 17 (43.6%) Other 4 (10.3%)	0.383
How was last baby delivered?	Vaginal 35 (84.6%) Forceps 1 (2.6%) Ventouse 1 (2.6%) Cesarean 1 (2.6%) Other 0 (0.0%)	Vaginal 33 (84.6%) Forceps 2 (5.2%) Ventouse 1 (2.6%) Cesarean 0 (0.0%) Other 1 (2.6%)	0.556

characteristics of the nondominant respondents to each exercise. There were no statistically significant differences in the personal characteristics of the two groups of nondominant respondents, indicating that any differences in results could not be attributed to differences in their underlying characteristics.

The results from the test of structural reliability for the nondominant respondents are presented in Table 3.

The simple comparison of models A and B indicates that there are some differences between the two models. In model B, the attributes that are statistically significant in influencing preferences are the level of continuity of care (CONT), decision making (DECIS), and the probability of being transferred during labor (TRANSFER). In model A, the only attribute that is statistically significant is the probability of being transferred during labor (TRANSFER). These findings suggest that those attributes in which the levels were varied attributes assumed a greater degree of importance in influencing preferences where the number of levels for these attributes was increased. The pooled data model (model C) is also presented in Table 3. This model reinforces the results obtained from the simple comparison of models A and B in that the differential slope coefficients for the continuity and decision-making attributes are statistically significant, indicating that the models from the two exercises are different.

DISCUSSION

The evidence from this paper relating to the structural reliability of DCEs in health technology assessment is essentially mixed. The proportions of respondents exhibiting dominant preferences were reasonably similar for both exercises. It was found that there were no statistically significant differences in the coefficient values for the constant term and the three common attributes in which the levels were not varied across the two models. These findings provide evidence of the structural reliability of DCEs. However, evidence against structural reliability was also found, since the relative importance of the two attributes in which the levels were varied increased as the number of levels increased. This finding has been reported elsewhere in previous applications of conjoint analysis in other disciplines (3). The results of this study suggest that there may have been a psychological effect in operation,

Table 3. Random Effects Probit Models

Attributes	Model A coefficient	95% Cl	Model B coefficient	95% Cl	Model C coefficient	95% Cl
Constant	−0.3007	−1.7101–1.1086	−0.2918	−1.6234–1.0958	−0.2849	−1.8891–1.0187
C	0.2757	−0.0055–0.5569	0.4148*	0.1452–0.6843	0.3281*	−0.0588–0.6214
L	0.2291	−0.0749–0.4149	0.2417	−0.1483–0.6716	0.2370	−0.3518–0.4282
P	−0.2952	−0.7254–0.1350	−0.3011	−0.8351–0.2329	−0.3218	−0.7552–0.0716
D	0.2170	−0.0711–0.5057	0.3740*	0.0552–0.7424	0.2864*	0.0351–0.6977
T	−0.4606*	−0.7914–0.0938	−0.4512*	−0.7203–0.1215	−0.4385*	−0.8822–0.0253
A_{ia}					0.1261	−0.0675–0.3120
$A_{ia}C_a$					0.2183*	0.0987–0.4214
$A_{ia}L_a$					0.2648	0.1758–0.6045
$A_{ia}P_a$					−0.1437	0.0351–0.6439
$A_{ia}D_a$					0.2190*	0.0643–0.4917
$A_{ia}T_a$					−0.0576	−0.3424–0.6970
N (data)	315		360		671	
N (groups)	36		40		54	
Chi^2	232		241		419	
p	<.001		<.001		<.001	

* = $p < .05$.

whereby nondominant respondents placed more importance upon specific attributes when the number of levels for these attributes was increased. However, it is very difficult to verify the true existence of a psychological effect within our study, since it was of a quantitative nature and was administered as a postal survey. It is important to investigate this phenomenon further within the context of DCEs conducted in health care and to use qualitative research techniques in attempting to understand how individuals actually make preferences between alternative scenarios in DCE studies.

If DCEs are to become an established tool for obtaining patient and public preferences in a healthcare setting, it is important that further research of both a qualitative and quantitative nature is conducted to assess the reproducibility (or otherwise) of these results in other DCEs in health care.

NOTES

[1] The hospital birth cohort was greater than the home birth cohort in the main study to allow for the potential for a lower response rate among the hospital birth group when compared with the home birth group, a finding observed in the pilot study.

[2] The hospital location was always the left-hand option within each choice, and hence a negative sign on the constant term would indicate an underlying preference for intrapartum care received in hospital.

REFERENCES

1. Bradley M. *Users manual for SPEED version 2.1 stated preference editor and designer*. The Hague: Hague Consulting Group; 1991.
2. Bryan S, Buxton M, Sheldon R, Grant A. The use of magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences. *Health Econ*. 1999;7:595-604.
3. Currim I, Weinberg C, Wittink D. The design of subscription programmes for a performing arts series: Issues in applying conjoint analysis. *Journal of Consumer Research*. 1981;8:67-75.
4. Propper C. The disutility of time spent on the United Kingdom's National Health Service waiting lists. *The Journal of Human Resources*. 1995;30:677-700.
5. Ratcliffe J, Buxton M. Patient's preferences regarding the process and outcomes of life saving technology. *Int J Technol Assess Health Care*. 1999;15:340-351.
6. Ryan M. Using conjoint analysis to take account of patient's preferences and go beyond health outcomes: An application to in-vitro fertilisation. *Soc Sci Med*. 1999;48:535-546.
7. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ*. 2000;320:1530-1533.
8. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. *Health Econ*. 1997;6:261-273.
9. Ryan M, McIntosh E, Dean T. Trade offs between location and waiting time in the provision of health care: The case of elective surgery on the Isle of Wight. *J Public Health Med*. 2000;22:202-210.
10. Van der Pol M, Cairns J. Establishing patient's preferences for blood transfusion support: An application of conjoint analysis. *Journal of Health Services Research and Policy*. 1997;3:70-76.

https://doi.org/10.1017/S0266462302108130 Published online by Cambridge University Press