ELICITING PREFERENCES FOR HERNIA REPAIR ~ A COMPARISON OF STRENGTH OF PREFERENCE MODELLING AND DISCRETE CHOICE EXPERIMENTS

Emma McIntosh[1] and Mandy Ryan[2]

[1] Research Fellow, HERC, Institute of Health Sciences, University of Oxford, Old Road, Oxford, UK, OX3 7LF
Email: emma.mcintosh@ihs.ox.ac.uk
Tel: 01865-226634 Fax: 01865-226842

[2] Reader, HERU, Department of Public Health, University Medical Buildings Foresterhill, Aberdeen, UK, AB25 2ZD

Paper Presented to the Health Economists Study Group Meeting

Brunel University, July 3rd-5th 2002

Preliminary analysis/work in progress:
Please do not quote without permission from the authors

Acknowledgement: This study obtained funding from the Chief Executive of the Scottish Office (CSO reference number: C2G/4/2/16). The views in this paper are those of the authors.

Abstract

Rationale

An alternative approach to the elicitation of preferences is 'strength of preference' (SOP) modelling. Environmental economists note that where the goods or services being valued are highly emotive, standard dichotomous choices may leave respondents dissatisfied with the amount on information they are allowed to provide on their intensity of preference for alternatives. As a consequence, they advocate the use of the SOP model to value natural resource damage assessment. It has also been identified in the marketing field that such an approach is a better predictor of longer-term adaptations to new products with relatively unfamiliar attributes. Further to this, the closed ended WTP literature has reported that indifference and ambivalence may be major cause of item non-response in closed-ended WTP surveys. Hence, if strength of preference models, with their capacity to identify indifference, are able to distinguish between true non-response and indifference (through comparing response frequencies with the standard approach) and are better suited to valuing unfamiliar attributes such as those in health care, this has implications for improving the estimation efficiency of the indirect utility model and thereby improving the statistical significance of the resulting WTP estimates.

Empirical study: Eliciting preferences for hernia repair

This paper aims to compare a SOP with a discrete choice model for eliciting welfare values for use in health care. Whilst there has been some work carried out in this area there appears to be a vacuum in the research as to the advantages and disadvantages of the strength of preference approach to DCEs with regard to eliciting welfare estimates. By comparing the two approaches in a head-to-head manner where all else is equal, information will be produced as to the differences (if any) in the welfare estimates obtained. The models will be compared in terms of the WTP values, variance around the WTP values, statistical significance of coefficients, log-likelihood results and McFaddens R^2. Qualitative interviewing will also provide information on patient's perceptions of the different approaches in a bid to determine whether Mackenzie's observations of the advantages of the strength of preference approach in agricultural economics also apply to health care. The qualitative interviews will also provide an insight to patient's views on the SOP format as well as levels of indifference, ambivalence and yea saying.

Introduction

Background to study

A number of trials have compared laparoscopic repair (LR) and open repair (OR) ((Liem MSL et al. 1997) (Champault et al. 1994); (Lawrence K et al. 1995); (Kald A et al. 1997); (Heikkinen T et al. 1997); (Wellwood J et al. 1998); (MRC Laparoscopic Groin Hernia Trial Group 2001)). The results from these studies suggest that there are differences between the two types of repair in terms of a number of *attributes* including short-term health outcomes and process characteristics. Each type of repair offers both advantages and disadvantages in relation to the varying attributes. There are clear trade-offs occurring between quality of life, return to usual activities, recurrence rates, pain scores and cost, with each repair offering different levels of each. However there is a lack of evidence on the importance or 'value' patients place on each of these attributes. Hence, despite the trial evidence on the many outcomes, it is unclear which of the repairs provides the greatest welfare gain to patients. This study carried out a discrete choice experiment (DCE) to establish the value of the attributes of hernia repair such that the welfare gains resulting from each surgical method could be compared in a cost-benefit analysis (CBA) framework.

The focus of this HESG paper however is to explore the methodological work carried out within this DCE of hernia repair attributes. Within the study a strength of preference (SOP) approach was compared to a standard binary approach. The following section outlines the rationale behind this comparison. The remainder of the paper will report solely on the methodological comparison of the SOP and binary models.

Strength of preference modelling

An alternative approach to the elicitation of benefits using DCEs is an approach called 'strength of preference' (SOP) modelling. Environmental economists, (Swallow S K, Opaluch J J, & Weaver T F 2000) note that where the goods or services being valued are highly emotive, standard dichotomous choices may leave respondents dissatisfied with the amount on information they are allowed to provide on their *strength of preference* for alternatives. As a consequence, they advocate the use of the strength of preference model to value natural resource damage assessment.

Mathews and Johnson (Mathews K E & Johnson 1999) used a strength of preference (graded-pair) DCE to obtain WTP estimates for non-use stated preferences in salmon preservation policy. This approach was used so that intensity of preference for the proposed policy could be obtained as well as direction of preference. The main difference in the strength of preference elicitation as compared to the standard discrete choice approach is that instead of being asked to identify their preferred scenario, X or Y, individuals are asked to rate the *intensity* of preference for their preferred scenario.

Huber et al (1993) (Huber J et al. 1993) note in the marketing field the discrete choice format is a better predictor of short term market behaviour in familiar, repeat purchase situations, while graded pair (strength of preference) "*is a better predictor of longer term adaptations to new products with relatively unfamiliar attributes*". In light of this observation and given the unfamiliarity challenges with regard to valuing health care attributes it seems that such an approach may be better placed to value health care attributes compared to the standard discrete choice approach.

Mackenzie (1993) (Mackenzie J 1993) compared the informational efficiencies of contingent rating, contingent ranking and pairwise comparison models and concluded that the contingent rating method is hypothesized to be most efficient because contingent ratings convey information on preference *intensities* and can also represent respondent *indifference*. The argument behind the use of this technique is that it allows respondents to express preference intensities as well as preference order. In addition to this, (Mackenzie J1993) state that the additional information should improve the estimation efficiency of the indirect utility model, thereby improving the statistical significance of the resulting WTP estimates. Strength of preference or graded pairs can be seen as a way in which to combine the advantages of rating data with the advantages of choice data i.e. a trade-off is being made explicit yet at the same time, respondents are being given the opportunity to provide additional information on the strength of preference or magnitude of the response.

The axiom of indifference has received little attention within the discrete choice literature, previous studies in health care have concentrated on exploring completeness and stability (Ryan and San Miguel, working paper 2002, San Miguel PhD Thesis Aberdeen University; Ryan and Gerard, iHEA 2001) continuity (McIntosh E & Ryan M 2002; San Miguel PhD Thesis Aberdeen University) and transitivity (McIntosh E & Ryan M 2002). One advantage of the strength of preference approach is its ability to capture respondent indifference. Further work is also required to explore the *behavioural* responses of respondents to such 'intensity of preference' models to identify whether such models are eliciting preferences which are more representative or as Johnson *et al* (Johnson, Banzhaf, & Desvousges 2000) note whether such approaches are eliciting improved preference data for attributes that are relatively *unfamiliar* such as health care attributes.

In health care, one study has explored the strength of preference approach alongside the standard approach (Johnson, Banzhaf, & Desvousges2000). This study however was not designed to directly compare the welfare estimates from the two elicitation formats and was more concerned with whether combining two elicitation formats yielded more valid and robust estimates than one approach alone. Further to this the experimental designs of the two elicitation approaches differed. Hence, whilst there has been some work carried out in this area there appears to be a vacuum in the research as to the advantages and disadvantages of the strength of preference approach to DCEs with

regard to eliciting welfare estimates. By comparing the two approaches in a head-to-head manner where all else is equal, information will be produced as to the differences (if any) in the welfare estimates obtained. Qualitative interviewing will also provide information on patient's perceptions of the different approaches in a bid to determine whether Mackenzie's observations (Mackenzie J1993) of the advantages of the strength of preference approach in agricultural economics also apply to health care.

Study methods

The study was carried out at two UK centres – London and Glasgow. Ethics approval for the study was obtained from the relevant ethics committees at each of the centres. The following stages were carried out in implementing the research and each of these stages will be described in the following sections:

Stages of the hernia study

- establishing the attributes and levels for use in the DCE
- using design software to obtain an orthogonal design matrix
- devising a covering letter and information sheet to send to patients along with the questionnaires
- piloting the questionnaires using clinic patients at both centres using verbal protocol analysis to obtain detailed information on preferences for informing the main survey instrument
- analysing the pilot study and making any amendments
- administering the main postal survey instrument to patients and analysing the resulting data
- devising welfare estimates for both the strength of preference and standard formats
- estimating the welfare shift from open to laparoscopic scenarios

Establishing the attributes and levels

The attributes and levels for the study were devised using own experience from trial data (MRC Laparoscopic Groin Hernia Trial Group2001), literature searching of the other open v's laparoscopic trials ((Liem MSL, van der Graaf Y, van Steensel CJ, Boelhouwer, Clevers, Meijer WS, & et al1997) (Champault, Marachly g, Rizk N, Lauroy J, & Boutelier1994); (Lawrence K, McWhinnie D, Goodwin A, Doll, Gordon A, & Gray A et al1995); (Kald A, Andberg, Carlsson, Park PO, & Smedh K1997); (Heikkinen T, Haukipuro K, Leppala J, & Hulkko A1997); (Wellwood J, Sculpher MJ, Stoker D, Nichols GJ, Geddes C, & Whitehead A et al1998) including a recent systematic review (Vale L et al. 2000) and consensus meetings with clinical collaborators. The attributes and levels outlined had to be representative of the main 'trade-offs' between laparoscopic and open groin hernia repair. In order to obtain welfare estimates, a payment vehicle was also included in the DCE.

Designing questionnaire scenarios using design criteria

The main criteria adhered to in this study were orthogonality of design as well as ensuring orthogonality was obtained in the differences when pairing scenarios. Level balance, the requirement that the levels of an attribute occur with equal frequency, was also adhered to (Huber J & Zwerina K 1996; Louviere J, Hensher DA, & Swait J 2000). Design software (SPEED, Hague Consulting) was used to identify an orthogonal matrix of scenarios in order to recover the 'main effects' [1]. Consistency checks were included using dominance criteria whereby two choice sets were included where all attribute levels were greater (or no worse) on all attribute levels and so the utility maximising choice was obvious.

Patient information

A covering letter and information sheet were devised for the patients. The information sheet was a requirement from the ethics committees however was also perceived to be useful in reminding patients of the *hypothetical* nature of exercise they were about to undertake. Previous studies have identified a phenomenon whereby people who were being surveyed with a stated preference (SP) questionnaire after having experienced the service interpreted the SP questionnaire as an evaluative instrument of the *actual* service they had experienced. Hence people were seen to be making attempts to find the scenario closest to their experience and ending up confused if it wasn't there. Since the hernia patients being sampled in this study were also people who had already experienced the service it was important to try to avoid this phenomenon. Hence, the information sheet provided an added opportunity to restate the hypothetical nature of the exercise. In a bid to maximise the response rates, the covering letters to patients were sent on the headed notepaper of the hospital they had attended and were addressed from their respective surgeon.

Pilot study

The standard procedure used by contingent valuation researchers is to apply a two step experimental structure to any study, whereby a pretest or 'pilot' is performed to obtain preliminary welfare results before performing the full welfare experiment. Such an approach was used in this study. The sample for the pilot study were obtained by attending follow-up hernia clinics run by the two surgeons in London and Glasgow. A think aloud technique 'verbal protocol analysis' (VPA), was used in the pilot study to elicit qualitative information on perceptions of the discrete choice task and use of the graded scale. Detailed practical guidelines on using VPA were followed (Svenson O 1989). Verbal protocol analysis response sheets were designed to note the responses to each DCE question. As recommended by Svenson (1989) (Svenson O1989) the method used for coding the VPAs was

founded on the representation system devised by one of the researchers (EM) based on a system that was felt to reflect the research problems being addressed, namely explore the SOP approach. Finally a marginal willingness to pay (WTP) question was included in the pilot study to obtain appropriate levels for the cost (value) attribute. This marginal approach represents a modification of the open-ended WTP approach (Shackley P & Donaldson 2001).

Analysis of results and devising welfare estimates for both the strength of preference and standard formats

The data will be analysed in two ways, using the strength of preference data in an 'ordered' fashion as well as 'imploding' the data re-interpreting it as binary choice. The strength of preference format or 'graded pairs' responses are ordinal ratings of utility differences between attribute level pairs. Any estimation strategy should therefore account for the discrete, ordinal nature of the response variable (Johnson, Banzhaf, & Desvousges2000). The appropriate econometric model for these data is ordered probit ((McKelvey R D & Zavoina W 1975). The utility difference for a profile pair *t*, dV^i_t, is as follows:

$$dV^i_t = A^i_{Rt} - A^i_{Lt} + Z + \varepsilon^i_t \qquad (1)$$

Where A^i_{Rt} and A^i_{Lt} are the indirect utilities associated with the right side (R) and left side (L) profiles respectively, and $\varepsilon^i_t = \varepsilon^i_{Rt} - \varepsilon^i_{Rt}$ is the associated disturbance term. where V(.) is the measurable component of utility estimated empirically and, $_{Rt}$ and $_{Lt}$ the levels of attributes in R and L (which include *cost* as an attribute if eliciting WTP), and Z socio-economic characteristics of the consultant that influence his/her utility and ε_j (j= L, R) reflects the unobservable factors in the individual's utility function The difference in indirect utility for commodity pair *t*, dV^i_t is specified as a simple linear function of attributes. In this model, we do not observe dV^i_t directly, instead we observe C^i_t which is a discrete rating category related to the unobserved dV^i_t of interest. The appropriate econometric model is therefore random effects ordered probit which incorporates both the discreteness and the natural ordering of the data. In such models, the data are stacked such that the preferred profile is on the right, making $dVit = V^i_{Rt} - V^i_{Lt} \geq 0$. This latter procedure assumes that subjects have no systematic preference for screen location (Johnson, Banzhaf, & Desvousges2000). The rating categories are recoded such that 0 indicates indifference and 3 indicates maximum difference.

[1] In the design literature in mathematical statistics, an 'effect' is a comparison of the means of the factor levels by means of orthogonal constraints (Louviere J, Hensher DA, & Swait J2000).

The imploded or 'suppressed' data are analysed using a random effects probit model readily available in LIMDEP (Greene, Greene, & Seaks 1995). These models are estimated within the random utility framework whereby the respondent will choose R (Right) over L (Left) if

$$V(A_R Z) + \varepsilon_B > V(A_L, Z) + \varepsilon_A \quad (2)$$

where V(.) is the measurable component of utility estimated empirically and, A_R and A_L the levels of attributes in R and L (which include *cost* as an attribute if eliciting WTP), and Z socio-economic characteristics of the consultant that influence his/her utility and ε_j (j= L, R) reflects the unobservable factors in the individual's utility function. Assuming a linear utility function V(A), the utility to be estimated in the choice questions in moving from scenario L to R is:

$$\Delta V=(\alpha_{0R} + \Sigma\alpha_{iB}A_{i_R} + Z + \varepsilon_R) - (\alpha_{0L} + \Sigma\alpha_{iL}A_{i_L} + Z + \varepsilon_L) \quad (3)$$

which can be simplified to:

$$\Delta V = (\alpha_{0R}-\alpha_{0L})+ \Sigma\beta_i A_i + e + u \quad (4)$$

where $\alpha_{0R}-\alpha_{0L}$ is the constant term in the model, reflecting the overall preference for B over A when there is no difference between the levels of attributes across scenarios, β_i, represent the parameters of the model to be estimated, and *e* and *u* are the unobservable error terms, where e is the error term due to differences amongst observations and u is the error term due to difference amongst respondents.

For both datasets confidence intervals for the welfare estimates were obtained by bootstrapping from the multivariate normal distribution of coefficients and their variance-covariance matrix (Greene, Greene, & Seaks1995). The 95% confidence intervals are the 2.5^{th} and 97.5^{th} percentile values from the bootstrapped distribution. The welfare estimates from each format will be tested for significant differences by testing whether the 95% confidence intervals overlap.

Pilot study results

The attributes and levels identified from the literature search are outlined in Table 1 below. Using the attributes and levels in Table 1, design software (SPEED, Hague Consulting Group) was used to identify an orthogonal matrix. One of the important design criteria outlined in the methods section is 'level balance', the requirement that the levels of an attribute occur with equal frequency. Table 1 shows that all the attributes have three levels except '*Type of anaesthetic'* which has two

levels. When the matrix of scenarios from the orthogonal design was identified using SPEED, it became apparent that there was an imbalance in the number of times 'general anaesthetic' appeared compared to 'local anaesthetic'. Of the 27 scenarios identified by SPEED, 'general anaesthetic' only appeared 9 times compared to 'local anaesthetic' which appeared 18. Hence in order to overcome this imbalance, it was decided to oversample the scenarios with 'general anaesthetic' levels. These additional scenarios were included in the design and paired off so as to obtain level balance. Upon pairing, the differences in the choice levels were then tested for collinearity using Pearsons and Spearmans correlation coefficients (to account for both discrete and continuous variables) in a bid to maintain orthogonality of the differences. The choices were then manipulated so as to reduce any overlap and this process was repeated a number of times to ensure minimal overlap. The resulting overlap was 25% on the total attribute levels, this is minimal as some overlap will always be present with factorial designs with as many attributes a this study. Care was taken to ensure that there was virtually no overlap on the 'cost' attribute as this is the key variable required for obtaining the welfare estimates and hence we need to obtain as much information from this as possible. Overlap of only 5% was present on the cost attribute.

Table 1 Hernia study attributes and levels

			Approximate estimates (based on literature review)	
	Attribute	**Levels**	**Open**	**Laparoscopic**
1	Type of **Anaesthetic**	General, Local	Local	General
2	Risk of **serious complications**	1 patient in every 1000 patients 5 patients in every 1000 patients 10 patients in every 1000 patients	1	10
3	**Number of days suffering post-operative pain**	3 days, 7 days, 14 days	14 days	3 days
4	**Long term persisting pain**	3%, 5%,13%	13%	3%
5	**Usual activities**	7 days, 14 days, 18 days	18 days	7 days
6	**Recurrance**	4%, 16%, 20%	20%	4%
	Payment vehicle			
7	**Loss of income to your household**	£100, £300, £500	£0	£350

Verbal protocol analysis pilot results

Eleven verbal protocol analyses pilot interviews were carried out in London's Whipps Cross hospital and Glasgow's Western General Infirmary. All interviews obtained charting information on the discrete choice specific questions. A charting exercise was used to record references to the three issues of interest. The verbal protocols were reviewed and comments related to these main themes were recorded. Only a brief summary of the VPA results are presented in this paper.

Nine out of eleven respondents provided responses to the marginal WTP question. The WTP values ranged from £160 to £3,000. The median value was £1,000 and the mean value was £1,018. There were some verbal reasoning provided for these answers, examples included:

> **ID L001** (Value: £1,000) "*All my savings for the better operation*"
> **ID L002** (Value: £1,500) " *1 month's salary*"
> **ID L007** (Value: £500) "*The same as three weeks wages – I'm self employed you see*"
> **ID G004** (Value: £160) "*Half a week's wages*"
> **ID G 005** (Value: £500) "*Half a month's salary, equivalent to the reduction in post-operative pain and usual activities"*

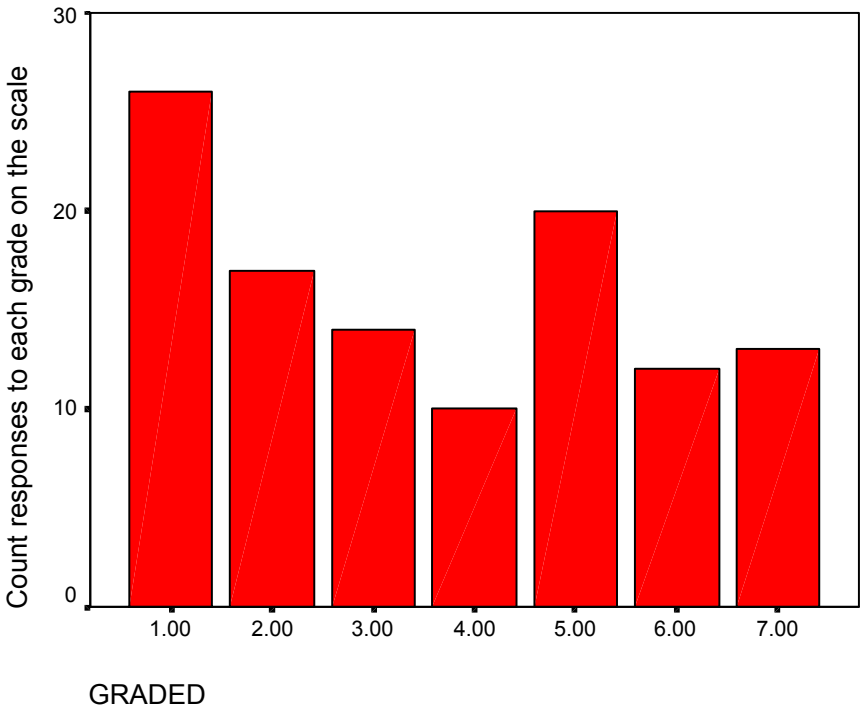
Whilst only a small sample size (n=9), the results imply that the original levels for the cost attribute were set too low. The mean level is £1,018 and the standard deviation is £844. Since 89% of the values fell between £0 and £1,500 it would seem that more realistic cost levels would be £500, £1,000 and £1,500.

Of the eleven respondents, nine showed evidence of using the response scale for the purpose of expressing intensity of preference. Of these nine people, four used the indifference option, sometimes more than once, and provided verbal reasoning for doing so. Examples of respondents using the indifference option (Grade 4) on the scale include ID L002 *: "this is a close one, difficult to choose – to me they are the same really"* also, choice 16, "*these are the same because of cost*". For ID L008, the indifference option was also used and the verbal notification provided the following reasoning "*This is really difficult because they are very similar – I would be happy with both equally*" and also in a later choice "*Pain days and usual activities are balancing out"* gave rise to the indifference option being chosen. For ID G002, the following reason was given for choosing indifference "*Local or general, no difference, would take the advice of the surgeon…these are equal*". ID G004 also choose an indifference option and his verbal notification was "*Difficult to choose, they are equal*" and again later on in another choice, simply "*Equal*". ID G005 used the indifference response twice, in choice 11 it was used to highlight a confusion in the questionnaire

which will be amended in the main survey, the verbal notification of this was "*These are equal*, *I'm confused because days post-op pain and days before usual activities are related but seem to act in the opposite*". This relates to the fact that days pain and usual activities are correlated and should therefore appear together but due to the design requirements for an orthogonal matrix, this didn't happen and the resulting design caused some confusion in some choice sets. This will be amended in the main survey. This same respondent also used indifference in choice 17 "*Pain and usual activities are bad for A but recurrence is bad for B, so they equal out*". Due to the small sample sizes in the pilot study and the main purpose being to obtain qualitative information on the key issues these data were not able to be analysed to explore the implications of using the graded scale on the welfare estimates, the main quantitative analysis will however explore this.

Figure 1 shows the extent to which the graded scale was used for responding and the range of the scale used. Of the 112 responses to the pilot graded scale in the WTP questionnaire, there was a broad spread of responses across the scale.

Figure 1 Histogram showing responses to graded scale



Two respondents showed clear evidence of their lack of use of the scale. ID L001 had a dominant preference for local anaesthetic and this resulted in their using the extremes of the scale each time (Grade 1 or 7). However, whilst the entire range of the scale was not used, one could argue that this respondent was using the scale correctly to express their intensity of preference for the dominant attribute. ID L005 was using the extremes of the scale and when prompted for their reasons for this

they replied: "(*I) ....will always use the extremes of the scale cause once I've made a decision, that's that !"*

MAIN STUDY

Sample & methods

The sample of patients for the main postal survey were identified from hospital records as having had a hernia repair in the past. In total, 600 patients were identified from existing databases, the majority of those had been involved in the MRC trials. These patients were sent a covering letter, information sheet and copy of the DCE-SOP questionnaire for self-completion and freepost reply. A reminder was sent after two weeks.

Following the results of the pilot study, a number of amendments were made to the WTP questionnaire, this included removing the attribute 'days to return to usual activities' as this combined with the attribute 'days pain following surgery' often created unrealistic combinations (and was highlighted in the verbal protocols). One basic (naturally occurring) consistency check was included as choice no.7. See Figure 2 for an example of the final questionnaire format.

Figure 2 Example of a choice set in the main questionnaire

Imaginary Operations

	A	B
Type of **anaesthetic**	**General**	**Local**
Risk of a **serious complication** giving rise to prolonged hospital stay	**0.1%**	**0.1%**
Number of days suffering **post-operative pain**	**3 days**	**7 days**
Cost of operation to you (£)	**£500**	**£1,000**
Chance of experiencing **long-term pain** up to 1 year after your operation	**5%**	**13%**
Chance of a hernia **recurrence**	**4%**	**16%**

1	2	3	4	5	6	7
A is much better	**A is somewhat better**	**A is slightly better**	**A & B are equal**	**B is slightly better**	**B is somewhat better**	**B is much better**

Main study ~ Results

245 respondents (3,185 possible choice responses) returned a completed questionnaire, this is a response rate of 41% so far however the questionnaires are still being returned (this is a preliminary analysis of the incomplete dataset). The characteristics of this group of respondents are shown in Appendix 1. The results of the consistency test included (based on dominance criteria) showed that 87.4% of the respondents were 'consistent' in choosing the expected scenario, 8.3% expressed an indifference preference and 4.4% were 'inconsistent'. There were 237 missing dependant 'ordered response' variables. Once removed, the dataset for analysis of the choice data contained 2948 responses. Figure 3 below shows the distribution of graded responses.

Figure 3 Distribution of graded responses

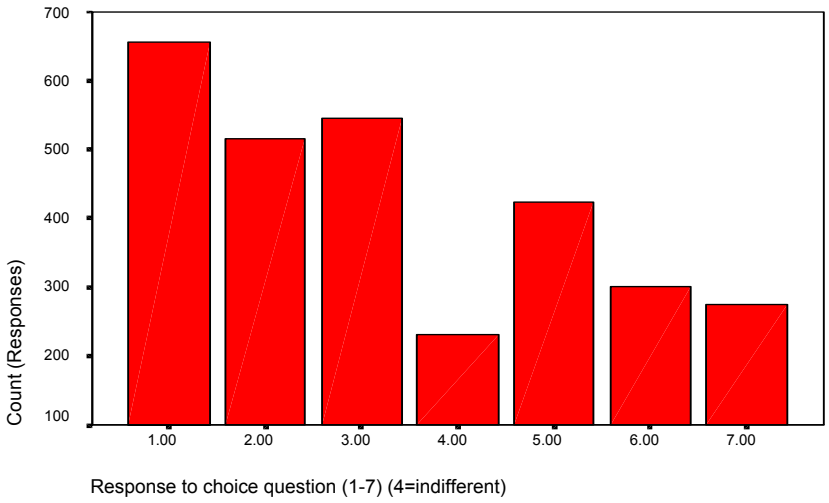


Figure 3 above shows that the graded response scale was used across the board, with little obvious sign of respondents preferring only the extremes of the scale (1 & 7). 29.6% of the responses were distributed at the extremes of the scale (response 1 or 7) whilst 232 (7.9%) of the responses were 'indifferent' between A and B. The results of the ordered probit model for the strength of preference model and the standard probit model for the 'suppressed' data are shown in Table 2 and Table 3 respectively.

Table 2 Random effects ordered probit model ~ strength of preference data

Variable	Coefficients (95% CIs)	SE	P	WTP (£) (95% CI's)
Type of anaesthetic (0=General, 1=Local)	**-0.0165** (-0.295, -0.1176)	**0.023699**	**0.000**	**£31.84** (£22.89, £32.94)
Risk of serious complications (%)	**-33.291** (-60.001, -27.946)	**5.011**	**0.000**	**£64,124.55*** (£46,096, £66,339)
Days in pain following surgery (days)	**-0.0601** (-0.0799, -0.05876)	**0.00353**	**0.000**	**£115.67** (£83.15, £119.66)
Cost (£)	**-0.000519** (-0.00072, -0.000501)	**0.000032**	**0.000**	/
Chance of long term pain up to 1 year (%)	**-0.04259** (-0.0632, -0.0322)	**0.00523**	**0.000**	**£82.04** (£58.97, £84.87)
Chance of recurrence (%)	**-0.0509** (-0.06462, -0.0455)	**0.00228**	**0.000**	**£98.18** (£70.58, £101.58)
Constant	**1.62** (1.394, 1.629)	0.0904	**0.000**	

Number of observations: 2948
Unbalanced panel: 234 individuals
Log likelihood function: -3208.99
Restricted log-likelihood: -3531.2
Chi squared: 644.43
Significance level: 0.000
McFadden's R^2: 0.09

*****_Extremely high value ~ possibly suggesting a lack of understanding of the risk attribute at such low levels or an interpretation of this attribute as the value of life._

Table 3 below shows the results of the data analysed as if binary choice data. To do this the data were re-coded to reflect a preference for A or B, those who were indifferent were randomly allocated to either A or B for estimation purposes only (sensitivity analysis will be carried out at a later stage to test this latter methodology).

Table 3 Random effects probit model ~ binary choice data

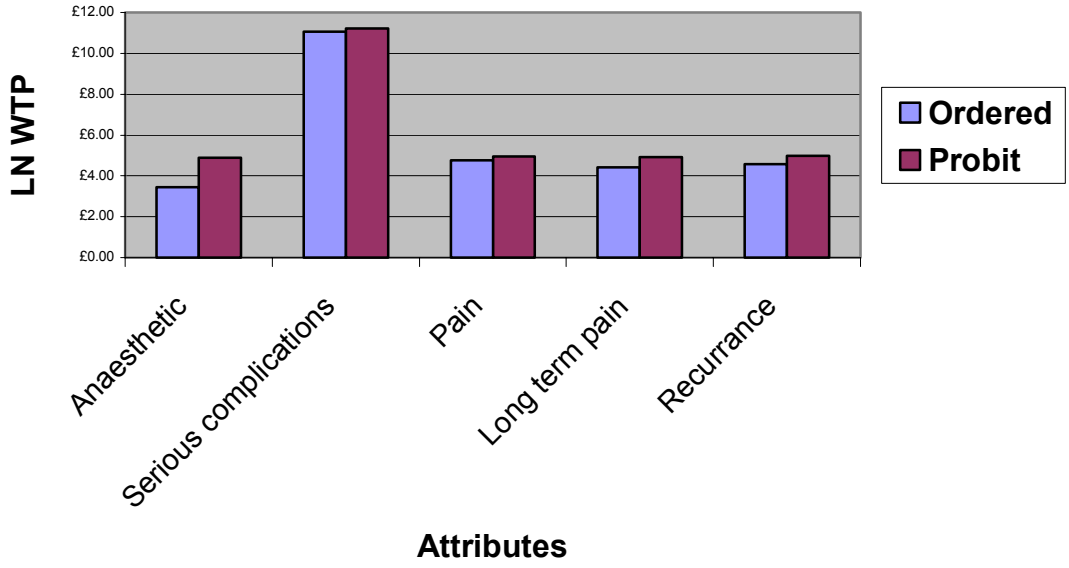
Variable	Coefficient (95% CIs)	SE	P	WTP (£) (95% CI's)
Type of anaesthetic (0=General, 1=Local)	**-0.070227** (-0.38503, -0.04265)	**0.028754**	**0.01**	**£135.13** (£114.49, £154.79)
Risk of serious complications (%)	**-38.9664** (-51.468, -26.477)	**5.4**	**0.000**	**£74,978.19*** (£63,525.95, £85,885.10)
Days in pain following surgery (days)	**-0.073105** (-0.09135, -0.07291)	**0.004988**	**0.000**	**£140.67** (£119.18, £161.13)
Cost (£)	**-0.0005197** (-0.000613, -0.00045)	**0.0000424**	**0.000**	/
Chance of long term pain up to 1 year (%)	**-0.0719496** (-0.1055, -0.06242)	**0.006953**	**0.000**	**£138.44** (£117.30, £158.58)
Chance of recurrence (%)	**-0.075011** (-0.084065, -0.0698)	**0.0023583**	**0.000**	**£144.34** (£122.29, £165.33)

Number of observations: 2948
Unbalanced panel: 234 individuals
Log likelihood function:-1512.27
Restricted log-likelihood: -1541.336
Chi squared: 58.122
Significance level: 0.000
McFadden's R^2: 0.02

***Extremely high value ~ possibly suggesting a lack of understanding of the risk attribute at such low levels or an interpretation of this attribute as the value of life.*

Figure 4 below compares the WTP (LN) responses from the two models. For comparison purposes the log if the WTP values has been used since it was not possible to compare the results on the same scale due to the large range of values (e.g. £13.84 - £74,978).

Figure 4 **Comparison of WTP (LN) values from strength of preference model and traditional binary choice model**



Discussion

The results presented in this paper present a preliminary analysis of a strength of preference model and a 'suppressed' binary choice model. The main outcome of interest are the welfare values obtained using each approach. The most obvious observation from comparing the WTP estimates in Tables 2 and 3 is that the welfare values derived from the SOP model (Table 2) are consistently *lower* than the suppressed model. Previous studies comparing choice probabilities using such models have reported similar results (Johnson, Banzhaf, & Desvousges2000; Johnson & Desvousges 1997). Three out of the five 95% confidence intervals on the welfare estimates do not overlap and hence could be said to be significantly different, these are: type of anaesthetic; long term pain; and recurrence. To see whether such welfare differences have any policy implications in this context we can use Table 1 to identify the main differences between open hernia repair and laparoscopic hernia repair. Estimating the welfare change moving from open to laparoscopic repair gives the following results: SOP model = -£573,425 ; imploded model = -£669,427. At an extra cost of £350 for the laparoscopic operation, these results imply that the additional cost is definitely not *worth* the benefits (from both models). Clearly these results are overwhelmingly dominated by the extremely large WTP value from reductions in risk of serious complications. One explanation is that people were interpreting this risk level as possibly 'risk of death' and hence these high values may reflect people's valuations of their own life. Further exploration into the interpretation of this attribute would be worthwhile. Excluding the risk attribute however reveals that all other attributes show clear net benefits moving from the open scenario to the laparoscopic scenario.

One explanation for the more general differences in the models is due to the inclusion of the indifference alternative in the SOP model. Previous work has suggested this may allow respondents to express valid preferences for indifference and ambivalence, previously ignored by binary format models. Although may rated pair studies have previously not employed utility theoretic estimation techniques, Roe *et al* (1996) (Roe B, Boyle KJ, & Teisl MF 1996) have shown how to link ratings to an underlying random utility framework. The extent to which the ordered probit model with its discrete/graded response variable gives rise to theoretically based welfare estimates is a matter for consideration. It may be that this model is not *theoretically* perfect but is more *behaviourally* accurate ~ which is best for using in economic evaluations? The lower welfare estimates produced in this paper by the SOP model imply that this model has allowed respondents to express preferences which are not as clear cut as binary choice assumes. However, the extent to which such SOP models actively *encourage* indifference and ambivalence is also worth considering.

The extent to which respondents in binary choice models exhibit yea-saying is also worth exploring within the context of this paper. There are two possible formats which could encourage yea-saying in choice experiments: absence of an opt-out alternative (such as those in unconditional model) and absence of an indifference option (such as those in SOP models). Neither of these models were unconditional due to the realism of such an alternative given the chronic health state being considered and the 'consumption' status of the sample. This combined with the fact that the attributes and levels were based on current practice and there were no realistic alternatives to treatment were reasons for using conditional choice models. Based on this, the significantly lower welfare estimates obtained using the SOP model may imply that some form of 'yea-saying' may be present in the binary choice format. Such 'yea saying' may be overcome by allowing graded responses of the type in SOP models such as 'A is somewhat better', 'A is slightly better' and 'A & B are equal'. Further consideration of this issue and the implications for the welfare estimates (& CBA results more generally) would be worthwhile.

Appendix 1 Characteristics of main sample

Characteristics		N (%)
Location	**Glasgow**	**134 (54.7%)**
	London	**111 (45.3%)**
Gender	**Female**	**24 (9.8%)**
	Male	**219 (89.4%)**
	Missing	*2 (0.8%)*
Age	**22-51**	**52 (25.3%)**
	52-72	**118 (48.2%)**
	73-87	**65 (26.5%)**
	Missing	*10 (4.1%)*
Qualifications	**None**	**84(34.3%)**
	O'Grade/GCSE	**36 (14.7%)**
	Higher/A'Level/SYS/OND	**16 (6.5%)**
	Diploma/HNC/HND	**14 (5.7%)**
	Undergraduate degree	**19 (7.8%)**
	Postgraduate degree	**15 (6.1%)**
	Other	**21 (8.6%)**
	Missing	*40(16.3%)*
Household Income	**Less than £9,999**	**62 (25.3%)**
	£10,000 - £14,999	**52 (21.2%)**
	£15,000 - £19,999	**25 (10.2%)**
	£20,000 - £24,999	**17 (6.9%)**
	£25,000 - £29,999	**18 (7.3%)**
	£30,000 - £34,999	**8 (3.3%)**
	£35,000 - £39,999	**9 (3.7%)**
	£40,000 - £44,999	**6 (2.4%)**
	£45,000 - £49,999	**5 (2%)**
	Greater than £50,000	**17 (6.9%)**
	Missing	*26 (10.6%)*
Ease of strength of preference questions?	**1 = Very difficult**	**4 (1.6%)**
	2	**2 (0.8%)**
	3	**13 (5.3%)**
	4	**24 (9.8%)**
	5=Moderate	**53 (21.6%)**
	6	**25 (10.2%)**
	7	**25 (10.2%)**
	8	**39 (15.9%)**
	9	**14 (5.7%)**
	10 = Very easy	**12 (4.9%)**
	Missing	*34 (13.9%)*
Type of hernia operation actually had?	**Open mesh**	**127 (51.8%)**
	Open non-mesh	**3 (1.2%)**
	Key hole (laparoscopic)	**76 (31%)**
	Don't remember	**32 (13.1)**
	Missing	*7 (2.9%)*

References

Champault, G. G., Marachly g, Rizk N, Lauroy J, & Boutelier, P. 1994, "Dans quelle conditions la traitment laparoscopique des hernies de l'aine est-il financierement rentable?", *J Chir (paris)*, vol. 131, pp. 483-487.

Greene, W. H., Greene, L. K., & Seaks, T. G. 1995, "Estimating the Functional Form of the Independent Variables in Probit Models", *Applied Economics*, vol. 27, pp. 193-196.

Heikkinen T, Haukipuro K, Leppala J, & Hulkko A 1997, "Total costs of laparoscopic and Lichtenstein inguinal hernia repair: a randomised prospective study", *Surg Laparosc Endosc*, vol. 7, pp. 1-5.

Huber J, Wittink DR, Fiedler, J. A., & Miller R 1993, "The effectiveness of alternative elicitation procedures in predicting choice", *Journal of Marketing Research*, vol. 30, pp. 105-114.

Huber J & Zwerina K 1996, "The importance of utility balance in efficient choice designs", *Journal of Marketing Research*, vol. 33, pp. 307-317.

Johnson, F. R., Banzhaf, M. R., & Desvousges, W. H. 2000, "Willingness to pay for improved respiratory and cardiovascular health: a multiple format stated-preference approach", *Health Economics*, vol. 9, pp. 295-317.

Johnson, F. R. & Desvousges, W. H. 1997, "Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs", *Journal of Environmental Economics and Management*, vol. 34, pp. 79-99.

Kald A, Andberg, B., Carlsson, P., Park PO, & Smedh K 1997, "Surgical outcome and cost-minimisation analyses of laparoscopic and open hernia repair: a randomised prospective trial with one year follow up", *European Journal of Surgery* no. 163, pp. 505-510.

Lawrence K, McWhinnie D, Goodwin A, Doll, H., Gordon A, & Gray A et al 1995, "Randomised controlled trial of laparoscopic versus open repair of inguinal hernia: early results", *British Medical Journal*, vol. 311, pp. 981-985.

Liem MSL, van der Graaf Y, van Steensel CJ, Boelhouwer, R. U., Clevers, G. J., Meijer WS, & et al 1997, "Comparison of conventional anterior surgery and laparoscopic surgery for inguinal hernia repair", *New England Journal of Medicine*, vol. 336, pp. 1541-1547.

Louviere J, Hensher DA, & Swait J 2000, *Stated Choice Methods: analysis and application*, 1 edn, Cambridge University Press, Cambridge.

Mackenzie J 1993, "A comparison of contingent preference models", *Americal Journal of Agricultural Economics*, vol. 75, pp. 593-603.

Mathews K E & Johnson, F. R. Dammed if you do and dammed if you don't: nonuse stated preferences in salmon-preservation policy. 1999.
Ref Type: Unpublished Work

McIntosh E & Ryan M 2002, "Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: impications of discontinuous preferences", *Journal of Economic Psychology (Forthcoming)*.

McKelvey R D & Zavoina W 1975, "A staistical model for the analysis of ordinal level dependent variables", *J.Math.Sociology*, vol. 4, pp. 103-120.

MRC Laparoscopic Groin Hernia Trial Group 2001, "Cost utility analysis of open versus laparoscopic groin hernia repair: results form a multicentre randomized clinical trial", *British Journal of Surgery*, vol. 88, pp. 653-661.

Roe B, Boyle KJ, & Teisl MF 1996, "Using conjoint analysis to derive estimates of compensating variation.", *J Environ Econ Manage*, vol. 31, pp. 145-159.

Ryan M, McIntosh E, & Shackley P 1999, "Using conjoint analysis to assess consumer preferences in primary care: an application to the patient health card", *Journal of Health Expectations*, vol. 1, no. 2, pp. 117-129.

Shackley P & Donaldson, C. Eliciting community preferences for health care using willingness to pay: is an incremental approach the way forward. 2001.
Ref Type: Unpublished Work

Svenson O 1989, "Eliciting and analysing verbal protocols in process studies of judgement and decision making," in *Process and structure in human decision making*, Svenson O & Montgomery H, eds., Wiley, Chichester, pp. 65-81.

Swallow S K, Opaluch J J, & Weaver T F 2000, "Strength-of-preference indicators and an ordered-response model for ordinarily dichotomous, discrete choice data", *J Environ Econ Manage* pp. 1-22.

Vale L, McCormack K, Scott N, & Grant A 2000, *Systematic review of the effectiveness and cost-effectiveness of laparoscopic versus open repair of inguinal hernia repair*, National Institute for Clinical Excellence.

Wellwood J, Sculpher MJ, Stoker D, Nichols GJ, Geddes C, & Whitehead A et al 1998, "Randomised controlled trial of laparoscopic versus open mesh repair for inguinal hernia: outcome and cost", *British Medical Journal*, vol. 317, pp. 103-110.