Eliciting patient and professional preferences for dental restorations

John Cairns[1], Ivar Espelid[2], & Jan Erik Askildsen[3].

1. HEB, Universitetet i Bergen & HERU, University of Aberdeen. 2. Institutt for klinisk odontologi, Universitetet i Oslo. 3. Institutt for økonomi, Universitetet i Bergen.

Introduction

The initial motivation for this study was to consider how different dental restorations might be valued for purposes of economic evaluation. This immediately raises the normative question whose values are appropriate, and the positive questions can the preferences of different groups be measured and do they differ. This study elicits the preferences of patients with respect to dental restorations and compares them with the views of dentists and dental assistants. The patients are Norwegian and Danish teenagers in a community dentistry setting. The dentists and the dental assistants are those caring for these patients. There are a number of methodological aspects to the study. Given that the patients in this context are teenagers rather than adults, whether or not they have well-defined preferences which can be measured is of particular interest. Discrete Choice Experiments and other methods of eliciting preferences have generally been used in health economics with adults. Also, the design facilitates an investigation of the impact of the complexity of the choices offered.

Discrete choice experiments are used to elicit patients' preferences for different dental restorations, and to elicit the recommendations of dental professionals with respect to different dental restorations. Other methods such as willingness-to-pay are less appropriate for both patients and for dental professionals. Dental care is free for these patients and even if it weren't questions regarding hypothetical willingness-to-pay for their dental health care would most appropriately be directed to their parents. With respect to the dental professionals interest focuses on what they think is good for their patients rather than for themselves. Standard gambles could be constructed with the risk of an inferior restoration rather than a risk of death as an outcome. However, as a practical matter the necessary interviews would make data collection on a similar scale infeasible. Quite apart from the difficulties associated with other methods a DCE appears particularly suited for the task of eliciting teenagers preferences. It appears natural to ask which of two restorations would one prefer or, in the case of dentists, which would one recommend.

The use of discrete choice methods to elicit preferences has been a major growth area in recent years (Ryan *et al*. 2001). Many of the published studies in the health field have been concerned with attributes of health care other than health outcomes. However, Szeinbach *et al*. (1999) and Hakim and Pattak (1999) elicited preferences over health states, and McKenzie *et al*. (2001) elicited preferences over asthma symptoms. The work reported here lies somewhere between the studies eliciting preferences over the process of care and those eliciting health state preferences.

The discrete choice experiment

The most relevant attributes of different dental restorations were identified following a small "pilot" study conducted with dentistry students. A general population survey (N=4530 adults) and a dentist survey (N=200) conducted by telephone in France, GB, Germany, Italy and Sweden at the end of 1998 were also consulted.

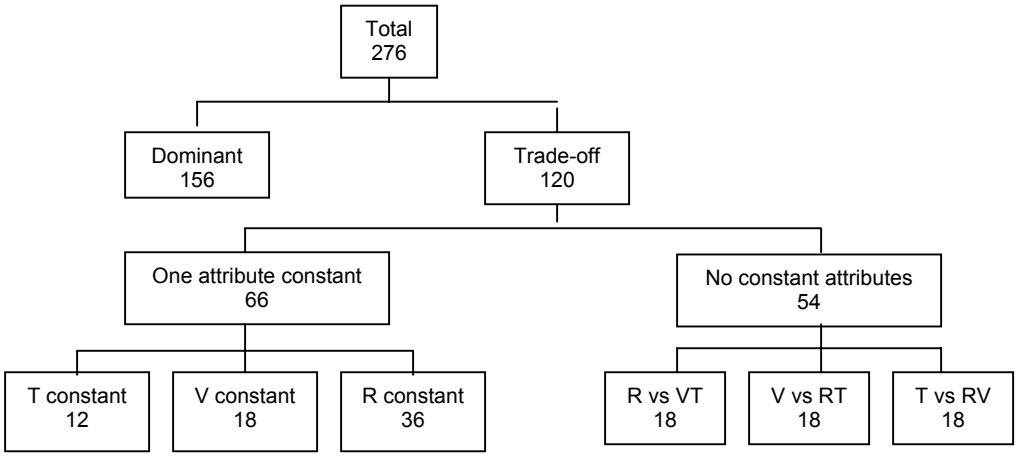
Three attributes were identified as being of greatest relevance: the expected duration of the restoration, that is the time before it is likely to need to be replaced; the visibility of the restoration; and the risk of an adverse reaction to the filling material. Two other attributes are, at first sight, strong candidates for inclusion, namely: the cost of the treatment; and some measure of its unpleasantness for the patient. However, neither is relevant in the context of the present study since dental care for this patient group is free, and the length of treatment, extent of drilling etc. will be the same for all restorations, all that differs are the properties of the material used for the restoration.

The questionnaires start with a preamble explaining the purpose of the study. The time until the restoration had to be replaced was set at either three, six, nine or twenty years. Three levels were defined with respect to the appearance of the restoration: not tooth coloured, highly visible; tooth coloured, but visible; and not visible. Photographs were provided which illustrated what was implied by the three descriptions of the visibility of the restoration. This has the advantage of making it more likely that all respondents both patients and dental professionals had a common understanding of the scenario. In a similar vein the preamble to the questionnaire presents information on the size and the nature of the risk of adverse reaction and a photograph of a patient with such a reaction was included. Because of the potential difficulties introduced by low probabilities only two levels of the risk of an adverse reaction were considered - no risk and

a very small risk. Also the preamble explained the small risk in terms of the number of cases expected per year in a particular city.

These attributes and levels give rise to 24 different scenarios from which 276 choice pairs can be constructed. The decision as to which choices to include was guided partly by the wish to test the axiom of transitivity and, to explore the impact of choice complexity in the teenage sample, and the recognition that not all choice pairs are equally informative. Given that there is a clear expectation regarding the preference ordering of levels (longer time to replacement is better, less visible is better, and lower risk of adverse reaction is better) a group of choices where there is a dominant option can be identified. There are 156 such choices where a participant's preferred option can be predicted with a high degree of confidence. The remaining 120 choices involve a trade-off between attributes so that *a priori* neither of the options is unambiguously to be preferred. These trade-off choices provide more information than ones containing a dominated option. In 66 of these choices one of the attributes is at the same level in both options; and in the remaining 54 no attribute is at the same level in both options.

Figure 1 Classification of potential choices



Dentists and dental assistants were presented with eighteen choices selected from these 54 ensuring that for each attribute there were six choices where it was traded against the other two attributes, and so as to provide a balanced distribution of attribute levels. Two questionnaires were constructed for patients, one "simple" made up of six choices with one attribute constant and another "complex" made up of six choices with no constant attributes. The choices were selected in order to reflect the distribution of the different types of choices in Figure 1, and the choices for the complex patient questionnaire were a subset of those presented to the dental

professionals. The pragmatic decision to restrict patients to six choices was based on the judgment that they might find the choices harder than the professionals, and that their willingness to participate might be relatively sensitive to the number of choices.

In order to facilitate the comparison of the preferences of dentists, dental assistants and patients it is necessary to minimise differences between the questions addressed to the three groups. Although it is appropriate to ask patients which restoration they would prefer, it is more appropriate to ask dentists which restoration they would recommend to their patients. Asking dentists which one they would prefer might encourage them to consider other aspects which are less relevant to patients such as, safety for the clinician, productivity and technical difficulty, and which are not explicitly addressed in the scenarios. A slightly different wording is chosen for dental assistants who are asked which restorations would be best suited for the patient. This avoids placing them in the unfamiliar position of making recommendations to patients but still doesn't encourage them to think of other aspects less relevant to patients.

Given that some pairs of scenarios may be viewed as equally good, patients are always offered three options: prefer A; prefer B; and "A and B equally good". Similarly dentists are given the option of not recommending either, and assistants are given the option of indicating that neither is best suited for the patient. As well as being more realistic, permitting indifference has the further advantage, that it provides a richer data set. This advantage comes at the cost of making the statistical analysis slightly more complex.

The choices in the simple patient questionnaire were selected to allow for a test of the axiom of transitivity (described below). In order to permit this test it is necessary to accept a very slightly unbalanced design with respect to attribute levels. In the case of the complex questionnaire it is impossible to construct a simple test of transitivity. However, this questionnaire can be balanced with respect to choice type and attribute levels. Patients are randomly allocated to the two questionnaires in order to investigate the impact of complexity of choices on patients' responses. A comparison of the results for the two patient groups can determine how much harder the choices are to make when all attributes vary. Since young patients are likely to have more difficulty answering these questions than adults they should be a good group in which to investigate the impact of increasing choice complexity. Increased complexity might make it more likely that participants express indifference or an inability to identify a preferred restoration,

and may increase the use of decision heuristics, such as optimising with respect to a single attribute.

In the simple questionnaire three linked choices were selected in order to test whether respondents are making selections consistent with or violating the axiom of transitivity. Each choice can be represented by a three-digit code. Thus 331 represents a scenario with time to replacement at level three (nine years), appearance at level three (not visible) and risk of adverse reaction at level one (very small risk). Let scenario 331 be represented by E; scenario 322 by F; and scenario 421 by G. Respondents are presented with three choices equivalent to E vs. F, F vs. G, and E vs. G. Ignoring the outcome "are equally good" there are four possible responses to the first two choices EF, EG, FF, and FG. A choice of E following EF is consistent with transitivity. Similarly a choice of G following FG is consistent with transitivity. Whereas, a choice of G following EF, or a choice of E following FG violates transitivity. Table 1 details the tests of transitivity allowing for expressions of indifference (I).

Table 1 Testing transitivity

Result of E vs. F and F vs. G	Result of E vs. G consistent with transitivity	Result of E vs. G violating transitivity
E F	E	G or I
E G	E, G or I	none
F F	E, G or I	none
F G	G	E or I
I F	E	G or I
I G	G	E or I
E I	E	G or I
F I	G	E or I
I I	I	E or γ

Figure 2 presents an example of one of the choices where the respondent is offered a potential trade of increased longevity of the restoration in exchange for increased visibility and risk. In practice the choices were in Norwegian or Danish.

Figure 2 Example of a discrete choice

Restoration A	**Restoration B**
This restoration will probably have to be replaced in twenty years. It is not tooth coloured and is highly visible. The risk of an adverse reaction is very small.	This restoration will probably have to be replaced in nine years. It is tooth coloured but is visible. There is no risk of an adverse reaction.

Statistical Analysis

The underlying model assumes that

$y^* = \boldsymbol{\beta}' \mathbf{x} + \boldsymbol{\varepsilon}$

Where y^* is the strength of preference for restoration B over restoration A. This cannot be observed but what is observed is the choice of restoration A, an expression of indifference, or the choice of restoration B.

$y = 0$ if $y^* \leq 0$

$= 1$ if $0 < y^* \leq \mu_1$

$= 2$ if $\mu_1 \leq y^*$

The strength of preference for B over A is assumed to depend on **x**: the difference in how long the restorations will last without being replaced (*duradiff*); the difference in the visibility of the restorations (*visidiff*); and the difference in the risk of an adverse reaction (*riskdiff*). The interpretation of the coefficients of these variables is not entirely straightforward. However, the expected signs are, in this case, clearly negative since as value of the difference variables increases it becomes less likely that restoration B will be chosen (that is, less likely we observe y = 2) and more likely that we observe y=0 (that is, restoration A is chosen). Increases in **x** might tend at first to increase the proportion where y = 1 but further increases are likely to lower the proportion reporting indifference. Given that it is not anticipated that there will be many indifferent responses and given that they are expected eventually to decline as **x** increases, the overall expectation is quite clearly that the coefficients of **x** (the difference variables) will be negative. However, interpretation of the coefficients remains difficult and thus it is more informative to use the coefficients to predict, for example, how the values of *duradiff* and *visidiff* affect the probability of a particular restoration being preferred.

Differences in preferences between respondents can be explored by including interaction terms between the individual characteristics and the difference variables. Data are available on the nationality, year of birth and gender of the patients, their dental health and their fear of dentists. It could be hypothesised that younger patients will attach less weight to differences in how long a restoration is expected to last, partly because they are less forward-looking and because compared to older patients the time when they will have to pay for dental care is further off. With respect to gender there is a popular perception that females attach a greater weight to their appearance than do males. Those respondents who are most afraid of dentists might be more sensitive to the risk of an adverse reaction and possibly duration in that they may particularly

wish to avoid future treatment. Finally patients who have experienced restorations might differ from those with none. The direction of any difference will probably depend on their previous experiences. It might lead them to attach less weight to risk because they regard it as smaller than treatment-naïve patients. With respect to duration if their experience has been favourable they might have fewer concerns than those yet to experience a restoration, on the other hand it might make them keener to avoid further restorations.

Given that the three possible outcomes can be viewed as ordered in terms of strength of preference for A over B (or for B over A), a regression model which takes into account the ordered nature of the dependent variable is required. One candidate is the ordered logit model (Long and Freese, 2001). The odds that an outcome is less than or equal m versus greater than m given $\mathbf{x}$ is Pr ($y \leq m|\mathbf{x}$) / Pr ($y > m|\mathbf{x}$). In the ordered logit model the log of these odds is equal to $\tau_m - \mathbf{x}\,\beta$. The assumption that the βs are the same for each value of m, is known as the proportional odds assumption or the parallel regression assumption. This assumption can be tested using a Wald test. If the assumption is rejected, an alternative is to estimate a generalised ordered logit model where the log of the odds is equal to $\tau_m - \mathbf{x}\,\beta_m$. That is, the generalised ordered logit doesn't require the βs to be the same.

While the three outcomes can be viewed as having an order, an alternative possibility is that the response "A and B are equally good" doesn't lie on a scale between prefer A and prefer B. In such a case a model for nominal outcomes, such as the multinomial logit model, would be more appropriate (Long and Freese, 2001).

Finally, since there are six observations per patient and eighteen observations per dental professional the observations cannot be assumed to be independent. The estimated standard errors need to be corrected to allow for this clustering of observations.

Data

The sample of dentists was made up of 22 from Norway and 20 from Denmark. These dentists were all participating in the broader dental study that initiated the enquiry into preferences over different restorations. There are also responses to the same questions from 65 dental assistants, 18 from Norway and 47 from Denmark. The assistants work in the practices of the dentists. In total there were 544 patient respondents but since one of these failed to answer any of the choice questions the effective sample size for this analysis is 543. Two hundred and fifty

three participants completed the simple questionnaire and 290 completed the complex questionnaire. There are 379 Norwegian participants and 164 Danish participants. Nine participants failed to answer three choice questions, two participants failed to answer two questions and five failed to answer one question.

For patients, in addition to their nationality and which questionnaire they completed, data are available on a number of characteristics of the respondents, namely: their gender; their year of birth; their fear of visiting the dentist; and their dental health measured by number of dental restorations. The distribution of participants by gender is fairly balanced (simple questionnaire - male 48.2 per cent, female 51.4 per cent, missing 0.4 per cent; complex questionnaire – male 47.6 per cent, female 52.4 per cent).

The distribution of participants by year of birth is shown in Table 2. Year of birth was collected in preference to date of birth in part to maintain anonymity of respondents. The original intention was to distinguish two groups a younger group (fourteen and fifteen year olds) and an older group (seventeen and eighteen year olds). However, the resulting data set is somewhat richer including thirteen and nineteen year olds. Indeed it is possible that the sample included a twelve year old and possibly a couple of twenty year olds.

Table 2 Respondents' year of birth

Year of birth	Simple Questionnaire	Complex Questionnaire	Total
1982		2	2
1983	3	8	11
1984	18	41	59
1985	89	76	165
1986	14	9	23
1987	27	36	63
1988	34	71	105
1989	63	41	104
1990	3	3	6
1991		1	1
missing	2	2	4
Total	253	290	543

Ninety-two respondents (16.9 per cent) had never had a restoration. Two hundred and seventy-two (50.1 per cent) had had between one and four restorations, and 177 (32.6 per cent) reported five or more restorations. Two respondents failed to answer the question. All of the participants answered the question with respect to fear of going to the dentist. Three hundred and seventy-

one (68.3 per cent) were not afraid of going to the dentist. One hundred and fifty-seven (28.9 per cent) were afraid to some extent and only 15 (2.8 per cent) were very afraid.

Because of the relatively small number of dentists and dental assistants no sub-group analysis other than by nationality was planned, and thus no other information was collected on individual characteristics from the dental professionals. A marked gender imbalance made this variable inappropriate for segmentation.

Results

Distribution of responses

The distribution of Prefer A and Prefer B responses are reported in Table 3. The "neither A nor B" responses are analysed in Table 4. For all but one of the choices (simple choice 3) an absolute majority of each group of respondents indicated a preference for one of the restorations. This suggests that the responses are not random. There is, however, evidence of quite widely differing preferences, particularly amongst patients, in that more than 70 per cent of respondents gave the same answer to only two out of the twelve choices. The choice generating greatest agreement amongst patients invited the respondents to accept a small risk of an adverse reaction and a highly visible restoration (rather than one which was not visible) in exchange for an increase of eleven years in the life of the restoration. This trade was only accepted by one in every eleven respondents.

Table 3 Distribution of choices

	Prefer restoration A				Prefer restoration B			
	Simple patients	Complex patients	Dentists	Dental Assistants	Simple patients	Complex patients	Dentists	Dental Assistants
1	58.2	51.4	85.7	76.9	26.7	30.7	9.5	15.4
2	34.3	17.7	2.4	12.3	57.8	76.7	95.2	84.6
3	45.0	55.2	90.5	80.0	43.0	34.1	7.1	16.9
4	22.7	85.2	73.8	70.8	68.9	9.2	14.3	24.6
5	22.0	59.2	85.7	69.2	68.8	28.7	4.8	23.1
6	32.7	24.4	69.0	35.4	61.8	68.9	28.6	60.0

There is agreement across complex patients, dentists and assistants with respect to five of the six choices, to the extent that a majority of each group chose the same restoration. However,

the size of this majority varies quite markedly in the case of choices 1 and 3. Dentists generally are in closer agreement with one another and patients tend to agree least with one another. In the case of choice 6 a majority of dentists favour restoration A whereas a majority of patients and of assistants choose restoration B.

In just over 10 per cent of the choices made by patients, dentists and dental assistants neither restoration was identified as superior. Complex patients and dental professionals are compared in Table 4 for the six choices they all faced. Patients are more likely to have expressed indifference on at least one occasion and overall a much higher proportion of their responses indicated indifference or an inability to choose. For these six choices there is only a slight difference between dentists and assistants – a higher proportion of assistants are never indifferent and the overall proportion of indifferent responses is slightly higher for dentists. However, when they are compared across the eighteen choices dentists express indifference more often – 7.7 percent of choices vs. 4.8 per cent of choices – and one third of dentists never express indifference as compared to two thirds of assistants. The proportion of responses indicating indifference was similar for the simple and complex patient samples. In both cases, about two thirds of respondents never expressed indifference possibly suggesting that the respondents to the complex choices did not have any greater difficulty answering than did the respondents to the simple choices.

Table 4 Extent of indifference

Number of ties (per cent)	Simple patients		Complex patients		Dentists		Dental Assistants	
Zero	166	(66.4)	187	(64.5)	29	(69.0)	50	(76.9)
One	46	(18.2)	61	(21.0)	12	(28.6)	10	(15.4)
Two	24	(9.5)	29	(10.0)	1	(2.4)	5	(7.7)
Three	10	(3.9)	8	(2.8)	0	(0.0)	0	(0.0)
Four	4	(1.6)	2	(0.7)	0	(0.0)	0	(0.0)
Five	0	(0.0)	1	(0.3)	0	(0.0)	0	(0.0)
Six	1	(0.4)	2	(0.7)	0	(0.0)	0	(0.0)
Total number of ties	146	(9.7)	168	(9.8)	14	(5.6)	20	(5.1)
Individuals with 1+ ties	85	(33.6)	103	(35.5)	13	(31.0)	15	(23.1)

Testing transitivity

Among the 253 respondents to the simple questionnaire there is some evidence of violations of the axiom of transitivity. As is shown in Table 5 nearly one in six of the respondents to the simple questionnaire failed the transitivity test. While there is some indication that violations of transitivity are influenced by the year of birth of the respondent (fewer in the oldest age group and more in the two youngest age groups) the relationship with year of birth is not statistically significant, possibly because the extent of violations is constant in the three middle age groups.

Table 5 Test of transitivity

Year of birth	Number	Consistent with transitivity	Violating transitivity
Before 1985	21	85.71	4.76
1985	89	83.15	14.61
1986	14	85.71	14.29
1987	27	85.19	14.81
1988	34	82.35	17.65
After 1988	66	78.79	21.21
All years	253	82.61	15.81

Note: 4 respondents could not be classified owing to missing data. And 2 were transitive but of unknown year of birth.

Traders and non-traders

Respondents can be described as non-traders when their answers are consistent with them regarding a particular attribute as determining the attractiveness of the restoration and the level of other attributes does not appear to influence their choice of restoration. The standard analysis treats the choices as if they were generated from respondents who are willing to trade. The appropriate treatment of these responses will depend on why individuals did not trade. If individuals would trade if offered more extreme trade-offs between the attributes, their individual marginal rates of substitution are not infinite, and there is no reason for excluding them from the statistical analysis. On the other hand, if when the attractiveness of one attribute is reduced, there is no change in the level of other attributes that can compensate the individual, the marginal rate of substitution is undefined and possibly infinite. It is less appropriate to include these individuals in the statistical analysis.

In order to provide a fairer comparison of the extent of trading and non-trading responses the dentists and the dental assistants have been assessed initially on the basis of the six choices

they faced in common with the complex patients. There are clear differences between patients, dentists and assistants in terms of trading and non-trading responses (Table 6). Traders form the largest single group in the case of patients, dentists and assistants. For dentists and assistants, the next largest groups are those whose choices are consistent with maximizing how long a restoration is expected to last, then those who consistently minimize the risk of an adverse reaction, and finally the smallest group is composed of those who consistently minimize visibility. Whereas, in the case of patients more than half of the non-traders minimized visibility, and maximizing how long the restoration would last was the least common response.

Table 6 Traders and non-traders

	Simple patients		Complex patients		Dentists		Dental Assistants	
	No.	Per cent	No.	Per cent	No.	Per cent	No.	Per cent
Traders	111	44.2	223	78.2	36	85.7	43	66.2
Non traders	140	55.8	62	21.8	6	14.3	22	33.8
Duration	20	8.0	9	3.2	5	11.9	13	20.0
Visibility	92	36.7	31	10.9	0	0	2	3.1
Risk	28	11.2	22	7.7	1	2.4	7	10.8

It is important to note that many of these non-traders might well have traded had they faced different choices. This can be illustrated by comparing these results with those where dentists and assistants are classified as traders and non-traders on the basis of their full set of eighteen choices. No dentists chose as if they were consistently optimising a particular attribute, however, however, five dentists still came close to choosing as if risk and visibility did not matter compared with duration of the restoration. Six assistants continued to make choices consistent with minimising risk and five with maximising duration. The reduction in the apparent scale of non-trading is partly explained by the greater opportunity to trade when offered eighteen choices, and it suggests that at least some of those who appeared to be non-traders are willing to trade when given a suitable opportunity.

There is a marked difference between the simple and complex questionnaires in terms of trading and non-trading responses. For both questionnaires the order of responses is the same. The largest single group is the traders, then those whose choices are consistent with minimizing visibility, then those whose choices are consistent with maximizing the length of time the restoration is expected to last, and finally the smallest group are those who consistently

minimize the risk of an adverse reaction. However, the scale of apparent non-trading is markedly different. At least part of the explanation for this is that in the simple questionnaire one attribute is constant in each choice and thus a wider range of responses are consistent with the appearance of non-trading than is the case with the complex questionnaire where all attribute levels vary. This difference between the simple and complex patient questionnaires implies that many of the respondents to the former questionnaire have the appearance of being non-traders rather than being unwilling to trade. Moreover, if apparent non-trading was a reflection of applying a decision heuristic in order to simplify the choice task the level of non-trading might have been expected to increase rather than decline with questionnaire complexity.

Regression analysis

The regression analysis which follows is in two parts: first, simple and complex patients are compared; and second, complex patients are compared with dentists and assistants. In both cases, ordered logit and generalised ordered logit results are presented, and then the predictions of the ordered logit model are compared with those of a multinomial logit model.

For the patient group as a whole the sign and the statistical significance of the coefficients of the three attributes estimated by ordered logit provides evidence of the existence of well-defined preferences. The coefficients are the expected negative sign for the simple and complex samples, and the combined sample. The only coefficient not to be statistically significant is that of *riskdiff* in the case of the simple sample. This may partly be explained by the less rich data with respect to *riskdiff*. In the simple questionnaire there is no difference with respect to risk level in three of the six choices, whereas duration differs in five choices and visibility differs in four choices. The only statistically significant difference in coefficients between the simple and complex samples is with respect to *riskdiff*. This is also attributable to the less rich data set generated by the simple choices.

A Wald test of the proportional odds assumption indicates that the assumption is not violated in the case of the simple sample but is violated in the complex sample and in the combined sample. The coefficients were re-estimated using a generalized ordered logit model. The results are not markedly different.

Table 7 Results for the ordered logit model

	Simple sample		Complex sample		Combined sample	
Ordered logit	Coefficient	P>|z|	Coefficient	P>|z|	Coefficient	P>|z|
duradiff	-0.192	0.001	-0.280	0.000	-0.297	0.000
visidiff	-0.657	0.000	-0.790	0.000	-0.755	0.000
riskdiff	-0.105	0.347	-0.752	0.000	-0.576	0.000
Wald χ^2 (3)	100.07		378.61		502.18	
Prob > χ^2	0.000		0.0000		0.0000	
Psuedo R^2	0.047		0.137		0.099	
Brant Prob > χ^2	0.172		0.002		0.014	
Generalised ordered logit	Coefficient	P>|z|	Coefficient	P>|z|	Coefficient	P>|z|
mleq1						
duradiff	-0.186	0.002	-0.231	0.000	-0.270	0.000
visidiff	-0.632	0.000	-0.755	0.000	-0.725	0.000
riskdiff	-0.128	0.260	-0.678	0.000	-0.546	0.000
constant	0.194	0.001	0.167	0.016	0.188	0.000
mleq2						
duradiff	-0.188	0.001	-0.304	0.000	-0.321	0.000
visidiff	-0.679	0.000	-0.812	0.000	-0.784	0.000
riskdiff	-0.041	0.735	-0.815	0.000	-0.613	0.000
constant	-0.289	0.000	-0.255	0.001	-0.274	0.000
Wald χ^2 (6)	136.43		455.82		611.50	
Prob > χ^2	0.000		0.000		0.000	
Psuedo R^2	0.049		0.141		0.100	
Sample size	1505		1717		3222	

A multinomial logit model was also estimated (which does not assume that the three outcomes are ordered). The results in terms of estimated coefficients are not straightforward to interpret in that they are expressed for two of the outcomes relative to the other outcome. However, the predicted probabilities of choosing restoration A (or of any other choice) based on the estimated model are much easier to interpret. The predicted probabilities for different combinations of *duradiff* and *visidiff* are shown in Table 8 for the ordered logit and multinomial logit models. The predicted probability of choosing restoration A *etc*. are very similar for the two models. The predicted probabilities of recommending restoration A over restoration B increase the more that restoration A is expected to outlast restoration B and the less visible restoration A is compared

with restoration B. These probabilities are closely mirrored when recommending B over A. The probabilities for A and B being equally good also vary systematically, as the differences between scenarios increase it becomes less likely that respondents will be indifferent.

The results of the sub-group analysis are summarised in Table 8. The coefficients of the interaction terms and their statistical significance are reported along with the LR test of the combined significance of the interaction terms. To interpret these coefficients remember that a positive coefficient on the interaction term combined with the negative coefficient of the difference variable would imply that that subgroup attaches relatively less weight to changes in that particular attribute than does the other subgroup other things being equal. As already noted those respondents who completed the simple question appear to give less weight to *riskdiff*.

When the respondents are classified as young (born after 1986) and not young (born up to and including 1986) there are no statistically significant differences between the coefficients of *duradiff*, *visidiff* and *riskdiff*. Female respondents attached a significantly smaller weight to the visibility of the restoration. Danish teenagers have different preferences from Norwegian teenagers. They attach less importance to visibility of the restoration and greater importance to risk of an adverse reaction. This effect is still significant when controlling for the nationality of the respondent. Those indicating that they fear going to the dentists attached a significantly smaller weight to the difference in risk of an adverse reaction. When the data are segmented into those with no fillings and those with one or more fillings there are no significant differences in the coefficients. However, when the data are segmented into those with 0-4 fillings and those with 5+ there is a significant difference between the coefficients of *riskdiff*. The latter group (those with poorer dental health) give less weight to the risk of an adverse reaction than do those with better dental health (and less experience of dental restorations).

Table 8 Significance of sub-group interaction terms

	duradiff		visidiff		Riskdiff		LR
Simple	0.0853	0.329	0.1325	0.110	0.6442	0.000	0.0000
Young	0.0910	0.092	0.0002	0.998	-0.0289	0.774	0.1859
Female	0.0031	0.954	0.1869	0.006	0.0071	0.944	0.0176
Danish	-0.0901	0.121	0.2445	0.000	-0.2977	0.006	0.0000
Dental fear	0.0139	0.815	0.0438	0.541	0.3793	0.001	0.0012
Dental health	-0.0649	0.266	-0.0797	0.268	0.2493	0.020	0.0013

Table 9 **Predicted probabilities from the ordered logit and multinomial logit models estimated on the combined patient sample**

	visidiff							
	-2		-1		1		2	
duradiff	**Predicted probability of choosing scenario A**							
-3	0.079	0.077	0.154	0.159	0.452	0.478	0.637	0.656
-2	0.104	0.103	0.197	0.205	0.526	0.546	0.703	0.709
-1	0.135	0.136	0.249	0.259	0.600	0.610	0.761	0.754
1	0.220	0.224	0.375	0.383	0.731	0.712	0.852	0.818
2	0.275	0.279	0.447	0.448	0.785	0.751	0.886	0.839
duradiff	**Predicted probability of A and B being equally good**							
-3	0.041	0.032	0.071	0.049	0.116	0.081	0.100	0.082
-2	0.052	0.042	0.084	0.062	0.113	0.091	0.087	0.088
-1	0.064	0.054	0.097	0.077	0.105	0.100	0.074	0.091
1	0.090	0.086	0.114	0.109	0.081	0.112	0.050	0.096
2	0.102	0.105	0.116	0.126	0.068	0.116	0.039	0.096
duradiff	**Predicted probability of choosing scenario B**							
-3	0.880	0.891	0.775	0.792	0.432	0.442	0.263	0.262
-2	0.844	0.855	0.719	0.733	0.361	0.363	0.210	0.203
-1	0.801	0.810	0.655	0.665	0.295	0.291	0.165	0.155
1	0.690	0.690	0.511	0.508	0.188	0.176	0.098	0.087
2	0.623	0.616	0.437	0.426	0.147	0.133	0.075	0.064

Table 10 presents the results for complex patients, dentists and assistants, first for the ordered logit model and then for the generalized ordered logit model. For all three groups the coefficients of the difference in duration, visibility and risk are the expected negative sign and are highly statistically significant. Dentists attach the greatest weight to differences in duration whereas patients attach the least weight to duration. The statistical significance of the differences in the coefficients between the three groups was assessed by LR tests of slope dummies estimated for the three samples pooled. Dentists attach a significantly higher weight to the difference in duration than do assistants, who in turn attach a significantly higher weight to the difference in duration than do patients. Assistants attach a significantly lower weight to differences in visibility than do either patients or dentists. The difference between patients and

dentists is not statistically significant. Finally the differences between the three groups with respect to risk of an adverse reaction are not statistically significant.

Table 10 Results for the ordered logit and generalised ordered logit models

	Patients		Dentists		Dental assistants	
Ordered logit	Coefficient	P>|z|	Coefficient	P>|z|	Coefficient	P>|z|
duradiff	-0.280	0.000	-1.468	0.000	-0.700	0.000
visidiff	-0.790	0.000	-0.907	0.000	-0.449	0.000
riskdiff	-0.752	0.000	-1.023	0.000	-0.756	0.000
Wald χ^2 (3)	378.61		112.15		170.65	
Prob > χ^2	0.0000		0.0000		0.0000	
Psuedo R^2	0.137		0.363		0.137	
Brant Prob > χ^2	0.002		0.015		0.520	
Generalised ordered logit	Coefficient	P>|z|	Coefficient	P>|z|	Coefficient	P>|z|
mleq1						
duradiff	-0.231	0.000	-1.447	0.000	-0.688	0.000
visidiff	-0.755	0.000	-0.833	0.000	-0.433	0.000
riskdiff	-0.678	0.000	-0.972	0.000	-0.741	0.000
Constant	0.167	0.016	0.073	0.369	0.017	0.735
Mleq2						
duradiff	-0.304	0.000	-1.492	0.000	-0.711	0.000
visidiff	-0.812	0.000	-1.006	0.000	-0.465	0.000
riskdiff	-0.815	0.000	-1.145	0.000	-0.782	0.000
Constant	-0.255	0.001	-0.549	0.000	-0.225	0.000
model χ^2 (6)	455.82		493.46		272.31	
Prob > χ^2	0.0000		0.0000		0.0000	
Psuedo R^2	0.141		0.369		0.138	
Sample size	1717		756		1169	

The Brant test indicates that the proportional odds or parallel regression assumption was violated in the case of patients and dentists but not in the dental assistant sample. However, the generalized ordered logit results in the bottom half of Table 10 are fairly similar to the ordered logit results for all three samples.

A multinomial logit model was also estimated for the three samples. Table 11 reports the probability of choosing restoration A predicted form the ordered logit and the multinomial logit model for different combinations of *duradiff* and *visidiff*. The first number in each pair is derived from the ordered logit model and the second number from the multinomial logit model. The predicted probabilities for A and B being considered equally good, and for choosing restoration B, are not shown in order to simplify the table and since they add little additional information. The predicted probabilities of choosing B are close to being a mirror image of those of choosing A since the probabilities of A and B being considered equally good are fairly low and vary over quite a limited range.

Table 11 Predicted probabilities for patients, dentists and assistants using the ordered logit and multinomial logit models

	visidiff							
	-2		**-1**		**1**		**2**	
duradiff	**Predicted probability of patient choosing scenario A**							
-2	0.107	0.120	0.209	0.239	0.562	0.613	0.738	0.772
-1	0.137	0.152	0.259	0.291	0.629	0.664	0.789	0.802
1	0.217	0.232	0.379	0.399	0.748	0.734	0.867	0.837
2	0.268	0.277	0.446	0.448	0.797	0.755	0.896	0.844
duradiff	**Predicted probability of dentist choosing scenario A**							
-2	0.006	0.003	0.015	0.010	0.087	0.082	0.191	0.199
-1	0.026	0.017	0.063	0.051	0.292	0.308	0.506	0.519
1	0.339	0.339	0.559	0.588	0.886	0.872	0.951	0.915
2	0.690	0.719	0.846	0.861	0.971	0.951	0.988	0.964
duradiff	**Predicted probability of dental assistant choosing scenario A**							
-2	0.076	0.074	0.115	0.113	0.241	0.244	0.332	0.338
-1	0.143	0.142	0.207	0.209	0.390	0.397	0.500	0.506
1	0.403	0.412	0.514	0.523	0.722	0.719	0.802	0.791
2	0.576	0.585	0.680	0.684	0.839	0.827	0.891	0.872

The impact of *duradiff* and *visidiff* on the choices made by patients, dentists and assistants is much clearer in Table 11. Consider how the likelihood of restoration A being chosen changes as the difference between the restorations in terms of visibility changes. For the case where *duradiff* equals –2 the probability of a patient choosing A rises by 0.631 (from 0.107 to 0.738 in the case of the ordered logit model). In contrast, the increase for dentists is much smaller

(0.185) and that for dental assistants is somewhere in between but closer to the value for dentists (0.256). Consider next the increase in the probability of choosing A as the relative longevity of the restoration rises. Holding *visidiff* at –2, a change form –2 to 2 for *duradiff* increases the likelihood of restoration A being chosen by 0.161 (patients), 0.684 (dentists) and 0.500 (assistants). These results highlight the greater importance attached to duration by dentists compared to visibility, and the greater importance for patients of visibility as compared to duration.

Discussion

This paper has two principal aims. The first is to establish whether or not it is possible to elicit preferences from young patients; and the second to compare these preferences with those of dental professionals. In addition the design of the analysis of patient preferences provides an opportunity to explore the impact of complexity of choice task.

The evidence suggests strongly that it is possible to elicit preferences for dental restorations from teenage patients. First, there was a very high rate of completion as evidenced by the limited missing data. Second, answers are clearly not random given that for each choice a majority express a preference for a particular restoration and the number of occasions on which they express indifference is relatively low. Furthermore, the regression results confirm the existence of empirical regularities. As the difference variables increase it becomes more likely that restoration B will be preferred. Third, while there is some evidence of violations of the axiom of transitivity the vast majority do not violate the axiom. There is some suggestion that younger patients are more prone to violate the axiom.

There is no evidence of an impact of increased complexity on either the proportion of choices where respondents did not indicate a preference or with respect to the proportion of respondents not indicating a preference on one or more occasions. There is a marked difference in the proportion of individuals answering in a manner consistent with optimising a particular attribute. However, the direction of the effect strongly suggests that much of the non-trading behaviour arises as a result of the more limited opportunities to appear to trade when one attribute is held constant. When choices become more complex (the level of all attributes varies) a much lower proportion of respondents answers as if they are optimising with respect to a particular attribute.

The direct comparison of patient choices with those of the dental professionals indicated broad agreement although the proportions selecting particular restorations differed quite substantially. The differences between patients, dentists and dental assistants are highlighted by the regression results and, in particular, by the associated predicted probabilities. The likelihood of a dentist favouring one restoration over another is very sensitive to the difference in the expected duration of the restorations, and much less sensitive to differences in visibility. The situation is reversed for patients, who display considerable sensitivity to differences in visibility and much less to differences in duration. The dental assistants lie somewhere in between these two groups but are closer to dentists than to patients in their views.

Thus it appears possible to elicit preferences from younger patients. This interesting finding suggests a raft of questions for further research: can this finding be generalised beyond dental care; at what age does it become difficult if not impossible to elicit preferences; how valid are these preferences; and how much more complex could the choices be without making it impossible to elicit meaningful preferences from this age group. Thus there is clearly a full and challenging agenda for future research.

Acknowledgements

Financial support was received for this project from Sosial-og Helsedir. The Health Economics Research Unit is core funded by the Chief Scientist Office of the Scottish Executive Health Department and HEB receives support from the Norwegian Research Council; however, the views expressed are those of the authors and not the funding bodies.

Address for correspondence

John Cairns, Health Economics Research Unit, University of Aberdeen, Foresterhill, ABERDEEN AB25 2ZD Email j.cairns@abdn.ac.uk

References

Hakim Z, Pattak DS. Modelling the EuroQol data: comparison of discrete choice conjoint and conditional preference modeling. *Health Econ* 1999;8:103-16.

Long JS, Freese J. *Regression models for categorical dependent variables using Stata*. Stata Press, College Station, Texas, 2001.

McKenzie L, Cairns JA, Osman *L.* Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. *Health Policy* 2001;57:193-204*.*

Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess* 2001;5(5).

Szeinbach SL, Barnes JH, McGhan WF Murawski MM, Corey R. Using conjoint analysis to evaluate health state preferences. *Drug Inf J* 1999;33:849-58.