European Journal of Orthodontics 22 (2000) 335–342

A comparison of health state utilities for dentofacial deformity as derived from patients and members of the general public

Susan J. Cunningham and Nigel P. Hunt
Orthodontic Department, Eastman Dental Institute, University College London, UK

SUMMARY The cost-utility approach is a method of economic evaluation, which assigns a ratio of cost to benefit, based on utility values of the health state in question. It allows efficient use of health care resources and is a useful method in that it permits comparison of a wide range of medical interventions, including those which are life saving and those that improve quality of life. This study obtained utility values for dentofacial deformity from orthognathic patients and members of the general public using three recognized methods—rating scale (RS), standard gamble (SG), and time trade-off (TTO). There were no significant differences between the utility values for the two groups of respondents. Method agreement between the TTO and the SG (the 'gold standard') was better than that between the RS and SG. In addition, the SG and TTO were found to have greater repeatability than the RS.

Introduction

Demand for health care has always exceeded the available resources. This raises the question of how to decide where the money should be most appropriately allocated. Economic evaluation is an accepted tool for the appraisal of health care programmes, and can be defined as the comparative analysis of alternative courses of action in terms of both their costs and consequences (Drummond *et al*., 1996). Evaluation is dependent upon the quality of underlying medical evidence and, because of this, clinical trials are increasingly viewed as a natural vehicle for economic analysis (Drummond and Davies, 1991). When 'rationing' of resources becomes necessary, some procedure has to be set to allow the most appropriate allocation. This was the basis for the introduction of cost-utility analysis (CUA) which assigns a ratio of cost to benefit and allows efficient use of resources in a manner that is considered consistent with justice.

Utility analysis

Most people would agree that quality of life is an important consideration in the allocation of resources and the emphasis in this area has increased dramatically over the last few years. CUA allows comparison of a wide range of different interventions, including those which are life saving and life enhancing (Gerard *et al*., 1993). Utility-based measures are usually expressed in terms of quality adjusted life years (QALYs). Health states are assigned a utility value between 0 and 1, where 0 is equivalent to death and 1 is equivalent to perfect health. QALYs are calculated by multiplying this utility value by the life expectancy. Information from QALYs, along with costs, can then be used to guide resource allocation. Utility values are therefore not the same as QALYs, but are used as a weighting factor to adjust the remaining life years for the quality of life that will be experienced (Bakker and van der Linden, 1995).

The main disadvantages of this method are that the interviews are labour intensive and, in addition, utility scores are not always easily interpreted. Despite these disadvantages, the technique remains popular and is likely to do so until a more appropriate method becomes available.

Existing measurement techniques

It has been widely assumed that the different measuring techniques produce comparable results, but this has not been investigated widely by practitioners of CUA. The following three methods are the most commonly used (Drummond *et al*., 1996).

Rating scale (RS). This is easily understood, but has no theoretical foundation. The measures from the RS tend to be referred to as values, rather than utilities because the latter are measured under conditions of uncertainty. Because the RS does not elicit valid cardinal utility measures, it should be used in conjunction with one of the other methods (SG or TTO).

The von Neumann and Morgenstern Standard Gamble. The standard gamble (SG) method is said to be valid because it is based on the utility theory for decisions under uncertainty (von Neumann and Morgenstern, 1953). This method has a number of advantages in that it reflects choice, the probabilities fulfil interval criteria, and the units are readily understood. In addition, it reflects risk, something that is involved in any medical intervention. It has been accepted as the 'gold standard' for many years.

Time Trade-Off. Torrance *et al.* (1972) originally developed the Time Trade-Off (TTO) for use in medical research because some respondents found the probabilities involved in the SG difficult to understand.

Cost utility and dentistry

There are few cost-utility studies in the field of dentistry. A recent Medline search found only 18 papers in this area published between 1980 and 1998. Fyffe and Kay (1992) assessed the average utility values for four different 'tooth states', which it was hypothesized would have different values. They found that the highest mean utility values were for the restored tooth and lowest values for the decayed and painful posterior tooth. Values were obtained from both dentists and members of the general public and, perhaps not surprisingly, dentists gave higher utility values.

Downer *et al*. (1997) used a convenience sample to elicit the public's perceptions of different oral cancer states (pre-cancer, small cancer, and large cancer). They used a SG questionnaire and found utility values of 0.92 for pre-cancer, 0.88 for stage 1 cancer, and 0.68 for stage 2 cancer. There were statistically significant differences between all three values with the order of magnitude in the expected direction.

The future in the field of orthodontics

The lack of papers in this field is somewhat surprising in view of the fact that this method is useful for interventions, which improve quality of life, as many dental interventions do. The utility approach is particularly appropriate for interventions such as orthodontics and orthognathic treatment where the main argument for undertaking treatment is improved quality of life. In the future, it seems likely that healthcare providers (such as the National Health Service or other insurance agencies) may ask for proof that procedures such as orthognathic treatment should be publicly funded. Utility analysis is one of the areas, which may then prove useful. The utility value is essentially a measure of how good or bad the individual perceives the health state in question. If patients with dentofacial deformity consistently provide high utility values (in the region of 0.95–1.0) then it may be questionable whether the service should be provided. However, if pre-treatment values are consistently lower than this and there is a significant change in the perceived utility value following treatment, then this provides a very good argument for the continued provision of the service.

This study aimed to obtain utility values for the pre-treatment health state (e.g. start of

treatment) for patients requesting combined orthodontic/maxillofacial surgery to correct dentofacial abnormalities.

Subjects and methods

The control group comprised a convenience sample of 55 adults (19 males and 36 females, mean age 24.0 years) from local colleges and businesses. All subjects were fit and healthy and none had any personal involvement in orthognathic procedures. The experimental group included 40 patients (13 males and 27 females, mean age 24.8 years) who presented with dentofacial deformities requesting orthognathic treatment in order to correct the problem. There was no significant difference in the gender or age distribution for the two groups.

The three methods already described (RS, SG, and TTO) were used to establish utility values for pre-treatment dentofacial appearance and function. One of the most important factors when using utility values derived from the general public is that they must understand fully the health state being described. The study aimed to achieve this using verbal explanations and black and white photographs of orthognathic patients. Photographic records of four patients were selected (male and female; Class II and Class III malocclusions) and laminated on card. The process of orthognathic treatment was also described to each respondent.

Rating scale

As described by Feeny and Torrance (1989) the RS was a vertical and calibrated visual analogue scale (0–100). The choice of a 0–100 dead-healthy scale was chosen such that the data obtained could be compared with a wide number of other medical interventions. Initially the ends of the scale were 'anchored'. This was achieved using two health state descriptions:

1. Perfect health state—respondents were 'forced' to assign perfect health to a scale value of 100.
2. Death—respondents were told that if death was the worst health state they could imagine, it should be placed at 0. However, if worse health states could be imagined (for example, being in constant pain), death was placed at an intermediate point (d) on the scale

The scale was then used to record a preference value (x) for the health state under test. If death was marked at a point (d) rather than at 0, the preference values of the health state took this into account using Equation 1.

$$\text{Value for health state} = x - d/100 - \text{d} \qquad \text{(eqn 1)}$$

Standard gamble

Subjects were asked to choose between two sets of health-related personal circumstances. In the first situation (Choice A), they were told to envisage that they would live with the dentofacial deformity for the remainder of their life. In the second situation (Choice B), they were asked to envisage a gamble situation where there were two possible outcomes, perfect health with no dentofacial deformity or immediate death (Figure 1). The probability of perfect health was decreased incrementally and the probability of death increased incrementally. The value at which the subject was not willing to take any further risk of immediate death was regarded as the indifference point and was recorded as the utility value. The more risk a respondent was prepared to accept, the lower the utility value.

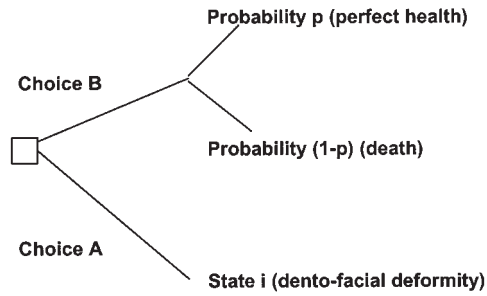

Figure 1 The standard gamble technique.

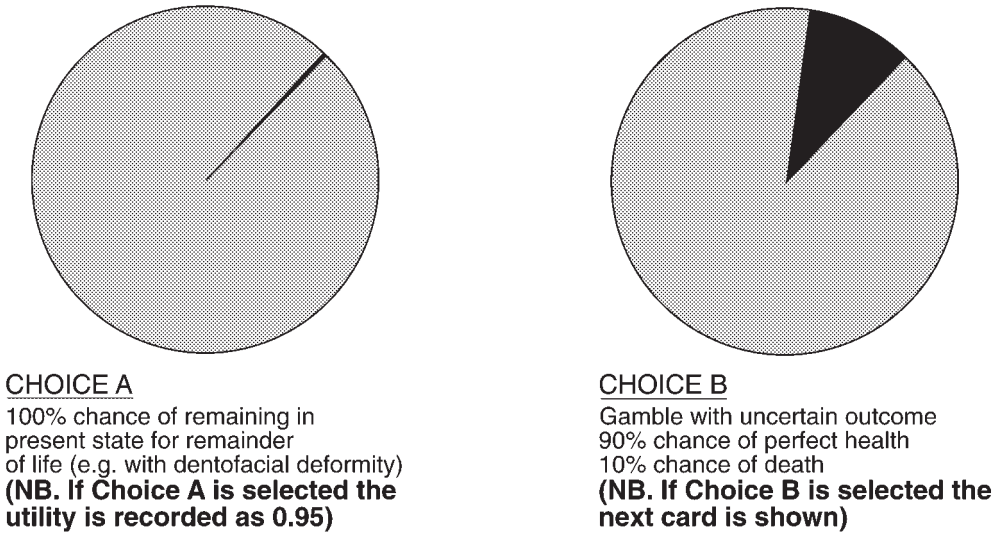

Figure 2 Example of card deck used in the SG.

The principles involved in the SG may be difficult for some respondents to understand and visual aids were used to make the concepts easier. A card deck was constructed using laminated diagrams in which colour-coded pie chart segments represented the probabilities (Figure 2). In order to avoid an anchoring bias, where values were constantly increased or decreased, a 'ping-pong' technique was used. This involved alternating values between high and low percentages (10/90, 90/10, 20/80, 80/20, etc.) and reduced the risk of respondents overshooting (i.e. over- or under-estimating the indifference point).

Time Trade-Off

The study assumed the remaining life expectancy to be 50 years, as the average age of the control group respondents was 24.8 years and that of the experimental group 24.0 years. The subjects were offered Choice A (State *i*), which was 50 years with dentofacial deformity or Choice B, which was a shorter life span, but in perfect health with no deformity (Figure 3). The time in the healthy state (time *x*) was varied until an indifference point was reached, at which the utility value was recorded. As with the SG, bias was reduced by alternating between long and short periods (for example, 45 years, followed by 5, 40, 10, 35, 15, 30, 20, 25). The more years a respondent was prepared to forego, the lower the utility value (Figure 4).

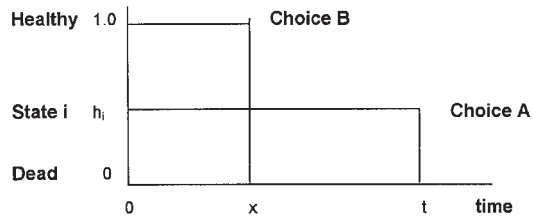

Figure 3 The TTO technique.

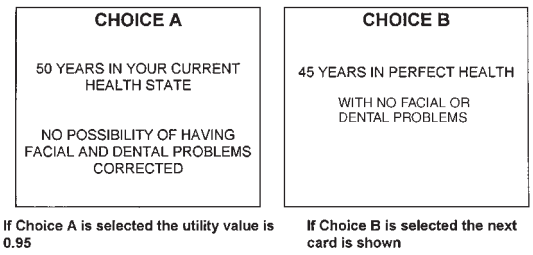

Figure 4 Example of card deck used in the TTO.

Table 1 Comparison of three methods (including test-retest data).

Group		Rating scale	Standard gamble	Time trade-off
Experimental	Median	57	0.85	0.75
	Range	10–93	0.05–0.95	0.05–0.95
	Mean (SD)	59.54 (19.76)	0.73 (0.24)	0.67 (0.25)
Control	Median	57	0.85	0.75
	Range	11–88	0.45–0.95	0.25–0.95
	Mean (SD)	54.25 (19.03)	0.79 (0.12)	0.72 (0.16)
Comparison of the two groups		0.37 (NS)	0.71 (NS)	0.27 (NS)
Test-retest	Product-moment correlation	0.68*	0.66*	0.84**
	Measure of agreement	0.15	0.09	0.09

*$P < 0.05$; **$P < 0.01$.

Repeatability of the method

In order to establish the repeatability of each of the three methods, 20 respondents underwent an identical interview 6 weeks after the first.

Statistical analysis

Data were analysed using the SPSS for Windows package (SPSS for Windows, SSPS International BV, Gorinchem, The Netherlands). Both mean and median utility values were calculated to avoid outlying results skewing the data. Comparison of utility values between the two groups used the Mann–Whitney test.

Test-retest repeatability and method agreement were determined using product-moment correlation coefficients, and the measure of agreement method recommended by Bland and Altman (1986).

Results

Table 1 shows the median and mean utility values (with standard deviations) for both groups from all three methods. The median value for the RS was 57 (with similar ranges) for both groups. Median values for the SG and the TTO were 0.85 and 0.75, respectively (for both experimental and control groups). However, the range and standard deviations were reduced for the control group. There were no significant differences between the mean values for the two groups for any of the three methods.

Table 2 Measure of agreement between the RS and TTO with the SG ('gold standard').

	Product-moment correlation	BSI reproducibility coefficient
SG versus RS	0.35**	0.43
SG versus TTO	0.70**	0.29

**$P < 0.01$.

Table 1 also includes the test-retest results. The product-moment correlation was significant for all three methods with the TTO giving the strongest correlation and greatest level of significance (0.84; $P < 0.01$). For the measure of agreement, the smaller the value, the greater the repeatability. The table illustrates similar levels of repeatability for the SG and TTO (0.09) with a higher value of 0.15 for the RS.

Table 2 illustrates the agreement between the values ascertained from the RS and the TTO when compared with the 'gold standard' (the SG). The product-moment correlation was significant for both methods, but the correlation was much lower when comparing the RS versus SG (0.35) than for the TTO versus SG (0.70). The BSI reproducibility coefficient for the SG versus RS was 0.43 and 0.29 for the TTO versus SG.

Discussion

One of the major benefits of the utility approach is that it produces a single measure that is amenable to statistical analysis. In addition, it is now an established method based on solid theory

(Torrance, 1987). The question of who should be asked to provide utility values has been the subject of much discussion. Torrance (1986) believed that informed members of the public should be used when public policy decisions were in question, the main factor being that respondents must be accurately informed. The investigation described above achieved this by use of both verbal descriptions and photographs. This study found that there were no significant differences between the utility values for the experimental and control groups. One of the underlying aspects to this study was that it was felt there may be a problem ensuring that members of the public were fully informed when determining utility values for dentofacial deformity and orthognathic treatment. Unlike other medical interventions, such as treatment for arthritis or angina, this is an area with which the majority of lay people may not be familiar. The fact that there were no significant differences between the two groups for any of the three methods of eliciting utility values suggests that members of the public understood the health state descriptions provided and that general public utility values may be considered a satisfactory source of data.

The three methods produced different values, as noted by Mooney (1994). The highest utility value was provided by the SG method and the lowest by the RS. The lower value from the RS reflects findings from previous research (Torrance, 1976). It seems probable that TTO values were lower than those from the SG because, although most respondents were quite willing to consider losing a number of years of their life when they had 50 years to consider, the extent of life/death gamble they were prepared to accept was more limited.

Interpretation of results obtained in utility studies is always difficult. Information is derived primarily by comparing the values with those from other studies. For example, the following utility values were obtained using a conventional TTO method (Torrance, 1987):

1. Some physical and role limitation with occasional pain (0.67).
2. Hospital dialysis (0.56).
3. Anxious and lonely much of the time (0.45).
4. Blind or deaf or dumb (0.39).
5. Hospital confinement (0.33).

Comparing these results with the present study where the mean utility value from the TTO method was 0.67, suggests that dentofacial deformity rates on a par with physical and role limitation with occasional pain, but was not considered as severe a problem as renal dialysis, being anxious or lonely much of the time, or being blind, deaf, or dumb.

A study of oral cancer (Downer *et al.*, 1997) using the conventional SG approach found utility values of:

1. Pre-cancer (assuming full recovery) (0.92).
2. Stage 1 oral cancer (also assuming recovery) (0.88).
3. Stage 2 oral cancer (assuming the cancer is life threatening) (0.68).

Again, when the results of the current study (the SG method gave a utility value of 0.73) are compared with these values, dentofacial deformity is considered a worse health state than either of the lesions associated with recovery, but not as great a problem as when the lesion is life threatening. Obviously, one point which must be borne in mind in this situation is that oral cancer tends to occur in an older age group than the group in this study and, as such, the subjects have to deal with dentofacial problems over much of their life, which may influence the results.

The difficulty of using the SG to look at health states not associated with mortality is in the selection of the 'worst health state'. However, this is quite a difficult concept for respondents to associate with and it may be that, although the SG is traditionally considered as the gold standard, the TTO may be more appropriate in this type of situation. Subjects in this study did not, however, appear to have difficulty in answering the SG probabilities.

This study used death as the worst health state in the SG and RS for two reasons:

1. It was felt that the oral cavity should be considered in the overall concept of a person's health

2. If the anchors used were not 'perfect health' and 'death', the utility values would have extremely limited use because they could not be compared with those for other health states which have used these traditional anchors.

If the SG is accepted as the 'gold standard', validity of the other techniques can be determined by comparison. The product-moment correlation for the SG versus TTO was stronger than for the SG versus RS (0.70 and 0.35, respectively). The BSI reproducibility coefficient provides an indication of the maximum difference likely to occur between two methods or, alternatively, it is the value below which the difference between paired results may be expected to lie with 95 per cent certainty. When the RS and TTO were compared with the gold standard, the BSI reproducibility coefficients were 0.43 and 0.29, respectively. This suggests that utility values derived by different methods are not interchangeable and illustrates the importance of quoting the method used, as well as the utility value obtained. Table 2 also shows the lack of comparability with the RS, which was found in other studies. Torrance (1976) believed that this may be due to the fact that no gamble is involved and suggested that the RS should always be combined with one of the other methods. Torrance (1987) suggested that the RS may be related to the other methods by means of a power curve rather than a linear function. The reason for the lack of agreement between the SG and TTO utility values has been discussed earlier.

The test-retest results in Table 1 show the highest product-moment correlation for the TTO (0.84) with slightly lower values for the other two methods. Similar measures of agreement were found for the SG and TTO (0.09 and 0.09) with a higher value (e.g. lower agreement) for the RS. Few studies report test-retest reliability results although Torrance (1987) quoted values of 0.63 to 0.80 for the TTO at 6 weeks, which is comparable with the present study. Torrance (1976) also reported values of 0.49, 0.53, and 0.62 for the RS, SG and TTO, respectively. However, as the retest was after a 1-year interval lower values would be expected.

The methods used were found to be acceptable to respondents. They appeared to understand the hypothetical situations and argued their utility choices with sound reasoning. There were no refusals to participate in the study and only one set of data for the control group could not be used as the interview was not completed. However, it must be borne in mind that the control group was a convenience sample.

Address for correspondence

S. J. Cunningham
Orthodontic Department
Eastman Dental Institute
256 Gray's Inn Road
London WC1X 8LD, UK

Acknowledgements

The authors would like to express their gratitude to Professor George Torrance of McMaster University, Ontario, and Dr Franco Sassi of the London School of Economics, whose advice in the early stages of this study was invaluable. The authors would also like to thank Drs Aviva Petrie and Ruth Holt for their statistical advice.

References

Bakker C, van der Linden S 1995 Health related utility measurement: an introduction. Journal of Rheumatology 22: 1197–1199

Bland J M, Altman D G 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 8: 307–310

Downer M C, Jullien J A, Speight P M 1997 An interim determination of health gain from oral cancer and precancer screening: 1. Obtaining health state utilities. Community Dentistry and Oral Epidemiology 14: 139–142

Drummond M F, Davies L 1991 Economic analysis alongside clinical trials. International Journal of Technology Assessment in Health Care 7: 561–573

Drummond M F, Stoddart G L, Torrance G W 1996 Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford

Feeny D H, Torrance G W 1989 Incorporating utility-based quality-of-life assessment measures in clinical trials. Two examples. Medical Care 27: S190–S204

Fyffe H E, Kay E J 1992 Assessment of dental health state utilities. Community Dentistry and Oral Epidemiology 20: 269–273

Gerard K, Dobson M, Hall J 1993 Framing and labeling effects in health descriptions: quality adjusted life years for treatment of breast cancer. Journal of Clinical Epidemiology 46: 77–84

Mooney G 1994 Key issues in health economics. Harvester Wheatsheaf, New York

Neumann J von, Morgenstern O 1953 Theory of games and economic behaviour, 2nd edn. Princeton University Press, New Jersey

Torrance G 1976 Social preferences for health states: an empirical evaluation of three measurement techniques. Socio-Economic Planning Sciences 10: 129–136

Torrance G W 1986 Measurement of health state utilities for economic appraisal. Journal of Health Economics 5: 1–30

Torrance G W 1987 Utility approach to measuring health-related quality of life. Journal of Chronic Disease 40: 593–600

Torrance G W, Thomas W H, Sackett D L 1972 A utility maximization model for evaluation of health care programs. Health Services Research 118–133