Choice of instrument for measuring the gain in utility from cochlear implantation

Garry R. Barton[1], A. Quentin Summerfield[2], David H. Marshall[2], Karen E. Bloor[1] on behalf
of the POCIA Collaboration[3]

[1] Department of Health Sciences and Clinical Evaluation, University of York, Heslington,
York Y01 5DD.

[2] MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD.

[3] The POCIA Collaboration involves clinicians in adult cochlear implant programmes in 14 NHS hospitals together with researchers at the MRC Institute of Hearing Research and the University of York. The lead surgeons and clinical coordinators of the programmes are: Mr J. Toner FRCS and Mrs C. McAnallen (City Hospital, Belfast), Mr D. Proops FRCS and Mr H. Cooper (Queen Elizabeth Hospital, Birmingham), Mr C. Raine and Mr S. Khan (Royal Infirmary, Bradford), Mr R. Evans and Mr. J. Joseph (Princess of Wales Hospital, Bridgend), Mr R. Gray and Mrs I. Court (Addenbrooke's Hospital, Cambridge), Mr A. Kerr and Dr. R. Barr-Hamilton (Royal Infirmary, Edinburgh), Mr J. Osborne and Ms M. Doran (Glan Clwyd Hospital, Bodelwyddan), Mr. R. Singh FRCS and Dr. A. Allen (Crosshouse Hospital, Kilmarnock), Mr J. Graham FRCS and Ms. W. Aleksy (Royal National Throat Nose and Ear Hospital, London), Mr A. Fitzgerald-O'Connor FRCS and Dr. T. Nunn (St. Thomas' Hospital, London), Professor R. Ramsden FRCS and Mr. M. O'Driscoll (Royal Infirmary, Manchester), Mr G.M. O'Donoghue FRCS, Mr. K.P. Gibbin FRCS and Mr. G. Armstrong-Bednall (University Hospital, Nottingham), Mr D. Chapman FRCS and Ms D. Gratton (Royal Hallamshire Hospital, Sheffield), and Mr P. Ashcroft FRCS, Mr M. Pringle FRCS, and Mrs J. Eyles (Institute of Sound and Vibration Research, Southampton).

Acknowledgements: Work supported by the Medical Research Council, Defeating Deafness—The Hearing Research Trust, and the National Lottery Charities Boar.

[Paper prepared for the Health Economics Study Group, Oxford, January 2001]

Work in progress – please do not cite without permission.

Abstract

Cochlear implantation (CI) is a high-cost low-volume intervention which is provided to about 200 adults with profound loss of hearing each year in the UK. Data are reported from a cohort of 147 adult patients who completed a battery of tests and questionnaires before and after CI. Both the EuroQoL instrument (EQ-5D) and the Health Utilities Index (HUI-3) showed a loss of utility associated with profound hearing impairment relative to population norms (EQ-5D: -0.05; HUI-3: -0.44). Both instruments showed a gain in utility after CI (EQ-5D: +0.03, 95% c.i. 0.00 to 0.06; HUI-3: +0.20, 95% c.i. 0.17 to 0.23). Tests of convergent validity showed that changes in utility measured by the HUI-3 were correlated with the primary purpose of improving hearing (measured by an increase in the accuracy of speech perception). In contrast, changes in utility measured by the EQ-5D were more correlated with an overall reduction in tinnitus, which is a secondary benefit of CI. Cost-utility ratios were £14,000 per QALY (HUI-3) and £93,000 per QALY (EQ-5D). The choice of instrument on which to base cost-effectiveness recommendations is important, therefore, given that interventions for impaired hearing must compete for resources with other interventions in health care. The results show that instruments which emphasise *health* (e.g. EQ-5D and SF-36) do not capture all of the aspects quality of life associated with a profound loss of hearing, nor all of the benefits of CI. Some of the attributes of the HUI-3 may be more suited to these tasks. Final judgement, however, must await the publication of the multi-attribute utility function for the HUI-3 in a peer reviewed journal, allowing public debate of its merits and de-merits.

Introduction

Twenty percent of the adult population of the UK suffers some degree of hearing impairment (Davis, 1995). For one in 25, the impairment is permanent, clinically significant, and affects both ears. Among people who are over the age of 70, the ratio is nearly 1 in 3. Hearing impairment is rarely the result of illness as conventionally construed, nor is it necessarily associated with ill health. Nonetheless, hearing-impaired people must compete for resources in health services when seeking treatments to alleviate their impairment. Economists argue that resource allocation should be based on principles of efficiency, where the weight given to health-care interventions reflects public preferences. In principle, cost-utility analysis achieves these goals. However, allocations would not maximise utility if utility measures did not fully reflect the preferences of individuals to endure sensory disabilities, relative to other health states.

We are addressing this issue by examining the properties of three widely-used instruments for measuring health related quality of life (HRQL): EuroQoL (EQ-5D), Health Utilities Index (HUI-3) and SF-36 (which can be used to derive the SF-6D). We assessed the construct validity (ability to detect hypothecated preferences) of all instruments to detect a loss of utility from being profoundly deaf, the responsiveness of instruments to cochlear implantation and the convergent validity (relationship between condition specific measures and utility instruments).

This paper has three parts. Part 1 compares the philosophies underpinning the development of the utility instruments and evaluates the techniques employed to derive and validate their measures of utility. Part 2 assesses the ability of the instruments to show a difference in utility between profoundly hearing impaired candidates for cochlear implantation and the general population, assesses the responsiveness of the instruments to cochlear implantation, and discusses the convergent validity of the instruments. Part 3 illustrates the impact on calculations of the cost-utility of cochlear implantation of adopting estimates of utility from the different instruments.

Part 1: Comparison of the derivation of the Health Utilities Index (HUI-3) and the EuroQoL (EQ-5D) instruments

There are two components of the task of estimating the utility of a person's health state: (i) Describing the state or profile of a person's health (i.e. choosing appropriate dimensions), and (ii) valuing the description of the state (i.e. deriving an index). This section outlines the content of the respective questionnaires and the approaches taken in estimating the tariffs (utility scores) for the EQ-5D and Health Utility Index Mark 3 (HUI-3). There is no objective criterion for judging the optimal way of undertaking these steps. It is of value, therefore, to compare the methods so that their robustness and suitability can be judged.

Derivation of the HUI-3

Choice of domains

The HUI-3 focuses on the capacity of individuals to function on the attributes of vision, hearing, speech (production), ambulation, dexterity, emotion, cognition, and pain. The philosophy can be characterised as "within-the-skin" because it focuses on the physical and emotional aspects of health status (Feeny *et al.* 1995). The within-the-skin approach is argued to result in a pure description of health status, uncontaminated by differential opportunity or preference (Feeny *et al.* 1995). The HUI philosophy is not intended to capture the effects of e.g. social interaction on health-related quality of life, which is argued to encompass individual preferences, choices, and opportunities and does not necessary reflect the level at which the individual is capable of functioning (Feeny *et al.* 1995).

The HUI-3 has evolved from the HUI-2 (Feeny *et al.* 1992), which in turn stemmed from a measure originally used for measuring quality of life in children with cancer. In developing the HUI-2 a literature review identified a set of 15 potentially relevant attributes. Eighty four parent and child pairs rated these attributes and the six most relevant attributes were chosen. They were sensory and communication ability (comprising vision, hearing and speech), happiness, self care, pain or discomfort, learning and school ability, and physical ability (Feeny *et al.* 1995). The HUI-3 overcame the overlap between self care and other attributes by replacing self care with a new dexterity attribute. Additionally, in order to increase precision, the sensation attribute was divided into separate attributes for vision, hearing, and speech.

Valuation of states

Although the HUI-3 is argued to have foundations in expected utility theory (Torrance *et al.* 1995), the latest "Multi-attribute Utility Function" (MAUF) is based on valuations of states using a visual-analogue scale (VAS). Values from the VAS were converted to estimates of scores that would have been obtained with the standard-gamble (SG) technique. This approach differs from the recommendations of Brazier *et al.* (1999) that SG and TTO values should be obtained directly, rather than being derived from VAS values. Furlong *et al.* (1998) justified the strategy on grounds of efficiency, and argued that it produced the same scores as the SG technique would have produced. The method is described in Furlong *et al.* (1998) and is summarised here with the permission of the authors. We have been asked to state that this reference should not be cited without the permission of the authors.

A graphical VAS was used to value health states. It took the form of a "feeling" thermometer on which a value of 100 represented the "best imaginable state" and zero the "worst state imaginable state". In making judgements, subjects were instructed to assume that states would be endured for their remaining life expectancy and to ignore effects on earning capacity. In total, 43 states were valued. This set was made up of three scaling anchor states: "Pits" (the lowest level on all eight attributes), dead, and perfect health; three methodological marker states (V2, H1, S1, A1, D1, E1, C1, P3; V2, H1, S1, A3, D1, E2, C1, P3; V2, H1, S1, A1, D1, E2, C3, P5); and 37 single-attribute states. A single-attribute state is an eight-attribute profile with one attribute at less than full function (level 2 or worse) and all other attributes at the highest level (level 1); e.g. "11511111". The HUI-3 contains 37 single-attribute states: 5 vision states, 5 hearing states, 4 speech states, 5 ambulation states, 5 dexterity states, 4 emotion states, 5 cognition states, and 4 pain states.

Two hundred and fifty six members of the general public aged 16 and over who were English speaking and non-institutionalised valued eight states using the feeling thermometer. Each respondent valued the three scaling anchors and the three methodological markers, plus two further states chosen randomly from the 37 single-attribute states.

Derivation of a predictive formula

The predictive formula for the HUI-3 is referred to as a "Multi-attribute Utility Function" (MAUF) by its originators. The derivation of the HUI-3 MAUF included the five steps. (i) Mean scores were calculated for each of the 43 states. (ii) The single-attribute states were normalised so that the "corner state" (the single-attribute state with one attribute at the lowest level) had a score of zero and the highest state (all attributes at level 1) had a score of unity. (iii) An end of scale bias adjustment (EOSBA) was included to compensate for the tendency for respondents to avoid values at the extremes of scales. (iv) The VAS values were converted to utilities using a power-curve estimate of the risk-aversion function, which varied according to whether the pit state or death was reported to be least desirable on the feeling thermometer. (v) Error-minimisation techniques were employed to determine the values of the constants a, b, and w_{ij} in a formula similar to the HUI-2 (Torrance et. al. 1996):

$$U = a + b\prod_{i=1}^{8} w_{ij}$$

where w_{ij} defines a set of weights for each of the eight attributes, i, and each level within an attribute, j. The result is a formula which converts a profile of levels on the eight attributes into a single value in the range from -0.35 to 1.00 that can be interpreted as a utility score.

Furlong *et al.* (1998) based the MAUF on a multiplicative function rather than an additive formula, as an additive function provided a poorer fit to the utility scores obtained for the corner states. The multiplicative function assumes that attributes are complements — the disutility of two attributes suffered together is less than the sum of the same two attributes suffered separately.

Validation of the formula

Two hundred and forty three new subjects used the SG technique to estimate the utilities of the same set of 43 health states, with each subject valuing a common set of six states — the three scaling anchors and the three methodological markers — plus two additional states chosen randomly from the 37 single-attribute states. The values were compared with the values predicted by the MAUF. On average, the utility predicted by the MAUF was 0.067 units higher than the standard-gamble scores. This correspondence, together with the results of other statistical evaluations, allowed the originators to conclude that the HUI-3 MAUF is a valid method for obtaining utility scores.

Derivation of the EQ-5D

Choice of domains

The EuroQoL philosophy focuses less on impairments and disabilities, and more on their consequences for daily living. The EuroQoL Group used their own expertise together with evidence from the literature to determine the dimensions of interest (Kind 1996). After modifications to the number of dimensions and levels within dimensions (Williams 1995; Brooks 1996) the current EuroQoL questionnaire (EQ-5D) focuses on five dimensions: physical functioning, self care, usual activities, pain/discomfort, and anxiety/depression. The EQ-5D was not intended to cover all aspects of health (Brooks 1996) and represents a compromise between comprehensiveness and simplicity in terms both of valuation and questionnaire design.

Valuation of states

In deriving weights for the EQ-5D, the time trade-off (TTO) technique was chosen in preference to the SG as it performed better in terms of internal consistency, sensitivity, and reliability (quality of data) (Williams 1995; Dolan *et al.* 1996a). As described by Dolan *et al.* (1996a), 6080 addresses were selected for sampling, of which 5324 were valid households. Interviews were conducted with one member of 3395 of these households (64%). Forty three states were valued by this sample of the general public. The states were spread across a range of mild, moderate, and severe levels of the dimensions to enable interaction effects to be tested. Each respondent ranked 15 health states (13 to be used in the TTO plus ‘11111’ and immediate death). Respondents were instructed to imagine that each state would last for 10 years and to rate it on a VAS scale where a value of 100 corresponded to the “best imaginable health state” and 0 corresponded to the “worst imaginable health state”. For those states ranked better than death respondents were asked to select the length of time in the ‘11111’ state that they regarded as equivalent to 10 years in the target health state. In the case of states worse than death, respondents chose between dying immediately and spending a length of time (x) in the target state followed by (10-x) years in the ‘11111’ state — the more time required in the ‘11111’ state (to compensate for the target state), the worse the state (Gudex, 1994).

Derivation of a predictive formula

Three criteria guided the selection of a formula for converting levels on the five dimensions into a composite value of utility (Dolan et. al. 1995): (i) Goodness of fit; i.e. how well the model explained the differences in the variations given to those states on which there is direct data. (ii) Parsimony; i.e. the simplicity of the model. (iii) Consistency; i.e. states that are logically worse must have lower predicted values. Data were analysed using a generalised least-squares regression technique. After testing many models[1] an additive formula was selected with the following form:

$$U = 1 - a(if\ any\ j < 1) - b(if\ any\ j = 3) - \sum_{i=1}^{5} w_{ij}$$

where (i) the weight given to the j^{th} level on the i^{th} dimension, w_{ij}, is independent of the weights given to other levels and other dimensions, (ii) a constant, a, is included if there is any departure from full health, and (iii) further constant, b, is included if any dimension is at its most severe level. The result is a formula which converts a profile of levels on the five dimensions into a single value in the range from -0.59 to 1 that can be interpreted as a utility score.

Validation of the formula

The average values given by respondents to the 42 states directly are numerically close to those produced by the model. The difference between the mean and estimated value exceeded 0.1 for only three states (Dolan *et al.* 1995).

Interim Conclusion

The relative theoretical merits of the TTO and SG techniques have been debated extensively (e.g. Torrance *et al.* 1995; Loomes *et al.* 1992; Dolan *et al.* 1996b). Other aspects of the methodology for deriving utility scores are also contentious. For example, Dolan *et al.* (1997) argued that VAS and SG scores are not related by a power function and Gudex *et al.* (1996) argued that utility scores depend on the length of time over which states are valued. Nonetheless, the EQ-5D and HUI-3 instruments are valuable. They enable a multi-dimensional description of health to be converted into a composite score which can be interpreted as a measure of utility. Therefore they can be used to compare the impact on health-related quality of life of different states of health and the relative merits of different interventions for those states. In the next section we discuss these issues with respect to cochlear implantation, a health-care intervention provided to some profoundly hearing impaired patients.

---

[1] Different numbers and types of interactions between dimensions were allowed for in models, however none of them improved the model significantly and many introduced inconsistencies into the estimated values.

Part 2: Loss of health-related quality of life associated with profound hearing impairment

Profound hearing impairment

Profound hearing impairment corresponds to the state that is colloquially described as being “deaf”. People with profound hearing impairment cannot detect sounds at many frequencies, while at other frequencies they can detect only those sounds that are so intense that they are painful. This loss of “dynamic range” — the difference between the intensities at which sounds can be detected and at which they become painful — limits the benefit that can be received from conventional acoustic amplifying hearing aids. The benefit is further limited because a profoundly impaired auditory system generates a distorted neural representation of those sounds which are audible. As a result, many people with profound hearing impairment cannot hear the sounds of speech with sufficient clarity to interpret what was said. They are unable to use the telephone. The quality of their own speech may deteriorate in the absence of veridical acoustic feedback about its intensity, rhythm, and frequency composition.

Profound hearing impairment is often also associated with tinnitus — “the conscious experience of a sound that originates in an involuntary manner in the head of its owner, or may appear to him to do so” (McFadden, 1982). Tinnitus is a symptom rather than a specifiable disease state. Both profound deafness and tinnitus often originate in damage to the inner ear (the cochlea) resulting from excessive noise, ototoxic agents, or genetic predisposition, either singly or in combination. The prevalence of tinnitus in the adult population ranges from about 10% of people who can recall an episode lasting five minutes or more, 5% for whom tinnitus creates material annoyance such as interfering with getting to sleep, and 0.5% for whom tinnitus has a severe effect on their ability to lead a normal life (Coles, 1997). Prevalence is materially higher among people with profound hearing impairments. About 80% report some degree of tinnitus, and about 20% report a disruptive degree. As yet, there is no objective technique for verifying tinnitus. Its loudness, psychological impact, and effect on daily living must be assessed by questionnaire.

Cochlear implantation

Cochlear implantation (CI) is a health technology which provides useful auditory sensations to people who are profoundly hearing impaired. This goal is achieved by implanting electrodes surgically in the cochlea so as to by-pass damaged parts of the peripheral auditory system and stimulate fibres of the auditory nerve directly with electrical signals. In the case of adults who acquired profound hearing impairment after learning spoken language, the primary aim of CI is to restore the ability to detect and understand speech. Some patients also find that CI attenuates, or eliminates, tinnitus. However, because other patients report that implantation induces or exacerbates tinnitus, CI is not provided as a treatment for tinnitus; rather, a reduction in annoyance due to tinnitus may be a worthwhile secondary benefit of implantation. A key aspect of the current study is to establish which instrument for measuring health-related quality of life best measures the loss of utility associated with profound deafness and the changes in utility associated with CI.

Case definition

Cases were adults (18 years of age or older), with profound hearing-impairment in both ears, whose impairment became profound after they had acquired spoken language. Their ages (Table 1) ranged from 18 to 79 years (mean 51 years, median 53 years). Prior to implantation, they scored less than 40% correct on a test of the ability to identify words in pre-recorded sentences without lipreading when optimally fitted with acoustic hearing aids. They underwent CI in one of 12 hospitals in the UK between January 1997 and December 1999, inclusive. They completed a battery of performance tests and questionnaires prior to implantation and again three and nine months after implantation. Data are reported from 147/200 cases, selected because they had completed all of a relevant sub-set of the tests and questionnaires at the pre-operative and nine-month stages.

Tests and questionnaires

Prior to implantation, patients provided baseline measures of accuracy of speech perception, annoyance due to tinnitus, and health-related quality of life. They also undertook other tests and questionnaires designed to predict the change in performance and in quality of life following CI. Nine months after implantation, patients repeated the baseline tests. Here we report results from a sub-set of one test and six questionnaires. Each test or questionnaire yields one or more composite scores. The abbreviation given to each composite score is placed in square brackets at the end of the description.

1. *Speech perception.* High-fidelity digital recordings of syntactically and lexically straightforward sentences were presented through a loudspeaker at a mean intensity typical of conversational speech. Patients reported as many words as they could. The number of content words reported correctly out of 100 was counted. *[Speech perception]*
2. *Tinnitus*. Sissons (1996) defined 13 questions which probe the subjective manifestations of tinnitus (e.g. "How loud would you say your tinnitus is?") and its psychological effects (e.g. "How often does your tinnitus make you feel helpless?"). Subjects responded by positioning a pointer on a visual-analogue scale whose ends were labelled appropriately for each question (e.g. "Extremely quiet"/"Extremely loud" or "Never"/"Always"). Factor analysis demonstrated that one factor underpinned responses to all questions. Responses were averaged and expressed as a single number in the range from 0 (no annoyance) to 100 (extreme annoyance). *[Tinnitus]*
3. *EuroQoL*. The EQ5D questionnaire and visual-analogue scale were presented. The number of patients who reported a problem on each dimension was counted. Additionally, reported levels of each dimension were converted to weights and the weights were combined additively in accordance with procedures defined by (Dolan et. al. 1995) to yield a value in the range from –0.59 to unity *[EQ5D-Q]*. The score on the visual-analogue scale was taken directly to yield a value in the range from zero to 100 *[EQ5D-VAS]*.
4. *SF36*. The questionnaire was presented in the form described by Ware (1993). Brazier *et al.* (1998) derived a preference based measure from the SF-36, the SF-6D, which has recently been updated (Brazier et. al. submitted). Results are presented here on the first algorithm, as the second is not yet published, and produce a separate VAS *[SF-6D VAS]* and SG *[SF-6D-SG]* scores.
5. *Health Utilities Index (HUI) Mark 2 and 3.* A single questionnaire allows both versions of the index to be calculated. It was used under license from the McMaster University

Center for Health Economics and Policy Analysis. With the agreement of the originators, the words "hearing aid" were replaced with "cochlear implant" in the version of the questionnaire presented post-operatively. Responses were scored in three ways. First, the number of patients was counted who placed themselves at a level below the highest level on each functional attribute. Second, a patient's level on each attribute was converted into a single-attribute utility score and subsequently converted into a utility score using a multiplicative formula, as described by Torrance et. al (1996) for HUI-2 and by Furlong et. al. (1998) for HUI-3. The resulting estimates of utility range from –0.03 to unity (HUI-2) and –0.35 to unity (HUI-3) *[HUI-2, HUI-3]*.

6. *Glasgow Health Status Inventory (GHSI).* The GHSI is a domain-specific measure of health status, composed of 18 items selected to be relevant to disorders of hearing (Robinson, Gatehouse, and Browning, 1996). Examples are "Is your self-confidence affected by any problem with your hearing?", "How often do you feel inclined to withdraw from social situations?", and "Which of the following statements best describes your view of the future?". Response are made on 5-point Likert scales, labelled for example (for the first question above): "Not at all affected", "Very slightly affected", "Slightly affected", "Moderately affected", and "Greatly affected". Each response is scored with an integer in the range 1 to 5. The 18 scores are summed and the sum is re-scaled linearly to a value in the range from 0 to100 *[GHSI]*.
7. *Glasgow Benefit Inventory (GBI).* The GBI provides a direct measure of the *change* in aspects of health and quality of life associated with an event, such as the onset of disease or an intervention intended to alleviate it. It is a partner to the GHSI insofar as each question in the GHSI has a corresponding question in the GBI. For example, with the event specified as cochlear implantation, questions take the form of "Since you received your cochlear implant, do you have more or less self-confidence?", "Since you received your cochlear implant, have you been more or less inclined to withdraw from social situations?", and "Since you received your cochlear implant, have you felt more or less optimistic about the future?". Responses are made on 5-point Likert scales that are labelled appropriately for each question. For the first question, the labels are: "Much more self-confidence", "More self-confidence", "No change", "Less self-confidence", and "Much less self-confidence". Each response is scored with an integer in the range from -2 to +2. The 18 scores are summed and the sum is re-scaled linearly to a value in the range from -100 to +100 *[GBI]*.

Our primary interest is in comparing the current version of the HUI (HUI-3) with the EuroQoL (EQ-5D) instrument. The HUI-2 is included for comparison with the HUI-3. The SF-6D is included because it is informative alongside the other generic measures, although we acknowledge that it is not based on values elicited from the general public and its algorithm has been updated (Brazier *et al.* submitted). The two domain-specific measures were included for comparison with the generic instruments.

Effects of cochlear implantation on measures of speech perception and tinnitus

Accuracy of speech reception increased from 3.8% correct (SD 9.3%) to 56.1% correct (SD 34.9%). The effect is significant (mean change 51.8%, 95% c.i. 46.0 to 57.6) and the effect size (1.46 = mean change divided by standard deviation of change) is large. Figure 1a is a plot which illustrates the effect. The 147 patients were ranked in increasing order of the size of their change in speech-perception score. Each patient was assigned a percentile rank. Change in score was plotted as the length of a vertical line against percentile rank. The mean change, and its 95% confidence limits, were plotted as a single data point and error bar on the right. Five percent of patients displayed scores that declined numerically; 8% displayed no change; 87% displayed scores that increased numerically.

Annoyance due to tinnitus fell from 23.9 (SD 22.6) to 16.8 (SD 21.8). The effect is significant (mean change -7.1, 95% c.i. -12.2 to -2.1), but the effect size is small (0.23). Figure 1b illustrates the effect in the same style as Figure 1a. Thirty eight percent of patients displayed scores that increased numerically; 5% displayed no change; 57% displayed scores that decreased numerically.

Thus, cochlear implantation had significant primary and secondary effects, although the secondary effect was small.

Effects of cochlear implantation on measures of utility

Methods of evaluation

The validity of an instrument should be judged in terms of what it is intended to measure. Utility instruments are intended to measure the value associated with different health states. Accordingly, we evaluated the EQ-5D, HUI-3, and SF-6D against the following three criteria:

Construct validity reflects the extent to which a measure is consistent with a plausible hypothesis (e.g. that possessing a profound sensory impairment reduces a person's health-related quality of life). We evaluated the instruments against this criterion by establishing whether the scores of patients differ from age-weighted population norms on one or more dimensions.

Responsiveness reflects the ability to measure changes in health-related quality of life[2] following either the onset of a condition or an intervention designed to alleviate it. We asked whether the scores of patients changed following CI both on composite measures of utility and on individual dimensions.

Convergent validity reflects the extent to which a measure correlates with another measure of the same concept. We assessed convergent validity by measuring the correlation between measures of utility and (i) measures of hearing (assessed by accuracy of speech perception) and of annoyance from tinnitus. We considered absolute measures obtained pre- and post-operatively, and also changes in measure between the pre- and post-operative states.

---

[2] It is acknowledged that utility based measures of HRQL are designed to measure the value that individuals attach to improvements post intervention and consequently that even though there may be a difference between pre and post intervention states this will not always change the utility level associated with that person's health. It is however our contention that, on average, these improvements are significant and consequently should increase the utility associated with a health state. Evidence of this, in an economic context, is provided by results from TTO questions which show a willingness of people to trade life years for improvements in hearing and receive a cochlear implantation (see later section).

There is the potential for circularity in the orientation to the issues that we have sketched here. An instrument that measures utility might normally be regarded as the gold standard. Therefore, if the instrument failed to meet some of our criteria, one would conclude either that the condition does not involve a loss of utility, or that the intervention does not restore some of what is lost. There is, however, independent evidence that profound hearing impairment is a condition that involves a loss of quality of life, and that implantation is an intervention that restores some of what is lost.

1. *Evidence that profound hearing impairment involves a loss of quality of life*. Most people with profound post-lingual hearing impairment seek help to alleviate the impact of their impairment. They claim that profound hearing impairment isolates them from their families and from society, may put their employment status at risk, increases tension within their families, makes them more dependent on other people for help in managing their lives, makes them depressed, anxious, and lacking in self-confidence, and reduces their enjoyment of life within and outside the home. These assertions are supported by data (Summerfield and Marshall, 1995) and by the domain-specific measures used in the present study.

   We also asked academic hearing researchers (mean age 38 years, S.D. 12 years) to value the state of profound hearing impairment using the TTO technique. Informants were asked to imagine that they were "profoundly hearing impaired", a state in which they would live until 75 years of age. They were then asked how many years they would be willing to give up in exchange for "completely normal hearing". The potential bias of this group is acknowledged. The loss of utility associated with a state of profound hearing impairment was 0.23 (95% c.i. 0.16-0.30).

2. *Evidence that cochlear implantation restores some of what is lost.* Patients seek and accept cochlear implants. Having received an implant, they contend that they would not be without it and seek immediate assistance should it develop a fault. The rate of elective non-use is low. Psycho-social benefits include increased confidence, reduced depression, and enhanced social and professional participation (Knutson *et al*, 1991; Harris, Anderson, and Novak, 1995; Summerfield and Marshall, 1995). Moreover when asked to rank their pre- and post-operative values of quality of life on a VAS, the mean scores improved by 0.23 post implantation (Summerfield and Marshall 1995).

   The same academic researchers were asked to imagine that they had "a cochlear implant in one ear" and that they would remain in that state until death at the age of 75. Again they were asked how many years of life they would be willing to give up to have "completely normal hearing". The mean loss of utility associated with a CI was 0.07 (95% c.i. 0.03-0.11). Thus, by comparison with the previous mean score, the informants were willing to trade life years equivalent to 0.16 on a utility scale to receive a cochlear implant.

We conclude that profound hearing impairment entails a loss of quality-of-life, though we acknowledge that it does not necessarily entail a loss of *health-related* quality of life, and that cochlear implantation restores some of what is lost. Accordingly, we feel justified in using deafness and implantation as a condition and an intervention with which to evaluate the construct validity, responsiveness and convergent validity of instruments that are intended to measure quality of life in ways that are relevant to decisions about resource allocation in systems of health care.

Measures on individual dimensions

EQ-5D

Kind *et al.* (1998, Table 2) reported the percentages of a representative sample of 3395 members the adult population of England, Scotland, and Wales who report a problem on each EQ-5D dimension. Results were presented for 10-year age-bands from 20-29 to ≥80 years. Age-weighted percentages were derived from these data to correspond to the age profile of the 147 patients. The percentages were converted to expected numbers out of 147 and are listed on the first line of Table 2. Subsequent lines contain the observed numbers of patients who reported a problem pre- and post-operatively with 95% confidence limits computed on the basis of binomial probabilities. The final lines assess the significance of the change in the observed numbers before and after implantation.

We hypothesised that profoundly deaf individuals would perform "usual activities" less well than the general population and that they would be more "anxious or depressed." On average, we expected these dimensions to show an improvement following implantation.

Results in table 2 show that pre-operatively significantly more patients reported a problem with anxiety/depression than expected from the age-adjusted norms. There were no other significant differences. The lack of difference in usual activities was unexpected and is addressed in the discussion. Post-operatively, more patients reported problems with mobility than expected from population norms. There were no other significant differences. Comparing the pre- and post-operative scores, implantation was associated with a significant change on one dimension: fewer patients reported a problem with anxiety/depression.

SF36

Jenkinson et al. (1996, Table 6) reported normative data for the dimensions of the SF36 broken down by age and gender for ages from 18 to 64 years. Age-weighted norms were derived from their data to correspond to the age profile of the 147 patients. In that calculation, it was necessary to apply norms for people aged 60 to 64 years to patients with a wide range of older ages (60 to 79 years). Therefore the age-weighted norms may over-estimate the health status of the patients. The age-weighted norms are listed in the first line of Table 3. Subsequent lines contain the observed values and their 95% confidence limits both pre-operative and post-operative. The final lines assess the significance of the change in the observed values before and after implantation.

We hypothesised that lower scores would be obtained from responses to questions in sections nine (mental health) and five (role limitations due to emotional problems) and possibly also section six (social functioning). We anticipated that cochlear implantation would alleviate some of these problems.

Pre-operatively patients reported poorer levels of physical functioning and social functioning than expected from the population norms, but higher levels of energy and general health. Post-operatively, patients reported poorer levels of physical functioning, role physical, and social functioning than expected from the population norms, but higher levels of energy and general health. Implantation was associated with a significant improvement in mental health.

HUI-2

Table 4 lists the numbers of patients who did not place themselves at the top *two* levels of functioning on the Sensation Attribute and the top level of functioning on each other attribute pre-operatively and post-operatively. There was a significant reduction in the number of patients reporting problems on the sensation attribute and a significant increase in the numbers reporting problems on the pain attribute. Table 6 lists single-attribute utility scores. There was a significant increase in score for the sensation attribute.

HUI-3

Population norms for the overall utility score and for each single-attribute utility score have been computed by Grootendorst *et al.* (2000) These norms were derived from the responses of 61,856 people in the 1990 Ontario Health Survey — a cross-sectional stratified randomised survey of the residents of the one province of Canada. The norms were not broken down by age. They are listed in the first line of Table 7 as expected values. The observed values, with 95% confidence limits, both pre- and post-operative are listed on subsequent lines. The final three lines assess the significance of the change in each single-attribute utility score between the pre- and post-operative measures.

We predicted that profound hearing impairment would be associated with lower levels than normal on three attributes: hearing, speech, and emotion. Table 5 lists the numbers of patients who did not place themselves at the top *three* levels of functioning on the Hearing Attribute and at the top level of functioning on each other attribute pre-operatively and post-operatively. There was a significant reduction in the numbers reporting problems on the Hearing Attribute.

Acknowledging the limitations of parametric statistics in situations where many scores are at ceiling, the observed utilities differ from the expected utilities for all attributes bar dexterity pre-operatively, and for all bar dexterity and cognition post-operatively. The existence of a large number of apparently significant differences may be a result of our not being able to compute age-weighted norms for purposes of comparison. The comparisons between the pre- and post-operative utilities of the patients, which are listed in the final lines of the table, are more informative, therefore. There are significant increases in the single attribute utility scores for hearing and speech, but for no other attribute.

Composite measures

Table 8 lists statistics (including mean and effect size) which describe the change in composite score obtained with each measure. A significant effect was displayed by five measures (HUI-II, HUI-III, EQ5D-Q, GHSI, GBI) and not by three (EQ5D-VAS, SF 6D-VAS, SF 6D-SG). Figure 2 plots, for each measure, the size of the change against each patient's percentile rank. Among the five instruments that displayed significant increases, large effects (>1) were displayed by the two domain-specific measures (GHSI, GBI), more modest effects (0.67-1) by the HUI-2 and HUI-3, and a small effect (<0.33) by EQ5D-Q.

Convergent Validity: Relationships with the primary and secondary purposes of the intervention

Given that utility scores measure a patient's quality of life, we expect a significant correlation between utility scores and improvements in hearing (measured by speech perception) and, to a lesser extent, with the reduction in annoyance due to tinnitus, both of which improved significantly following implantation.

Table 9 lists correlation coefficients computed between the measures of speech perception and tinnitus and each of the eight composite measures. Coefficients were calculated separately with the pre-operative scores, the post-operative scores, and the changes in score. The entries in the first two columns of the table show that there is no association between speech perception and annoyance due to tinnitus, nor between the changes in those measures that accompany implantation. Thus, although tinnitus and impaired speech perception may originate in a common cochlear pathology, their manifestations and the changes in them induced by cochlear implantation, are not related. This independence is convenient because it allows the correlation of other measures to speech perception and tinnitus to be assessed without variations in one outcome confounding variations in the other.

Given that a large number of coefficients were calculated, the criterion for significance was set conservatively at $p<0.01$. Pre-operatively, there was little inter-subject variability in speech-perception scores (the majority of patients scored 0%) and only one composite measure (HUI-2) showed a significant correlation. There was considerable variation in annoyance due to tinnitus which correlated significantly with each of the utility measures. Post-operatively, there was high inter-subject variability in speech perception which correlated significantly with the composite measure of utility from the HUI-2, HUI-3, SF 6D-SG and the SF 6D-VAS score. There was also high inter-subject variation in annoyance due to tinnitus which correlated significantly with the composite measures from each of the generic measures of health status. Four of the changes in composite measures were sensitive to the change in speech perception but not the change in annoyance due to tinnitus (HUI-2, HUI-3, GHSI, and GBI). The EQ5D-Q was more sensitive to the change in annoyance due to tinnitus than the change in speech perception. Two were sensitive to neither change (SF 6D-VAS, SF 6D-SG).

Interim Summary

Taking the results obtained with the three utility instruments together, the following picture emerges of the changes in self-reported health-status that accompany cochlear implantation in post-lingually deafened adults. Patients report that their capacity to hear speech improves and is accompanied by an increase in the intelligibility of their own speech (HUI-2, HUI-3). They suffer less anxiety and depression (EQ5D-Q) and mental health (SF-36). In addition, they report a net increase in pain according to one strategy for scoring the HUI-2. No single instrument displays all of these effects, however.

The instruments provide estimates of the gain in utility from CI that range from a non-significant gain of less than 0.01 (SF-6D-SG), to significant gains of 0.03 (EQ5D-Q), 0.14 (HUI-2), and 0.20 (HUI-3). Three of these utility measures are responsive to CI: the HUI-2, HUI-3, and EQ5D-Q. The responsiveness of the two versions of the HUI is associated with the improvement in speech perception that results from CI. The responsiveness of EQ5D-Q is associated with a reduction in annoyance due to tinnitus.

Part 3: Cost-utility ratios

The seven-fold ratio between the smallest and largest of the significant estimates of the gain in utility might not matter if the cost of providing CI and of sustaining implanted patients was very small. In fact, the cost is sufficiently high to mean that the smaller gains yield cost-utility ratios that fall outside the range considered to be competitive within the British health-care system. Thus, the choice among the instruments is important.

The itemised costs of providing CI to adults in five UK hospitals have been measured prospectively over the duration of the current study and retrospectively back to the inauguration of each programme. These data will permit estimates of cost-effectiveness and cost-utility to be based on average and marginal costs, and will allow economies of scope and scale to be identified. Results reported here are based on the average cost of providing a cochlear implant in the financial year 1998-9. Average costs are based on the resource use associated with providing cochlear implants (all implant programme staff, equipment, accommodation, indirect costs, in-patient costs, consumables and implant hardware), with salary costs based on levels at 1st April 1999. Future costs of rehabilitation and implant upgrades every six years have been included, in line with previous estimates Summerfield and Marshall (1995). All costs are discounted at 6% *per annum*.

The mean age of people receiving a CI in the UK is 51 years. The average life expectancy of a person receiving a CI has been estimated to be 30 years (Summerfield and Marshall, 2000). The median average cost of providing a cochlear implant, and further rehabilitation, over a 30 year period is £43,582. A small proportion of candidates for CI present with acoustic hearing aids. Potentially, these patients would continue to use acoustic hearing aids if they did not receive a cochlear implant. We estimate that the costs averted by CI are small, primarily because of the low cost of the acoustic hearing aids dispensed within the NHS. The average cost of supplying current NHS analogue aids has been estimated to be £90, with digital devices likely to cost an extra £250 (Taylor and Paiseley 2000). The upper estimate of the costs averted by cochlear implantation, assuming that every patient would otherwise have received bilateral digital hearing aids that were renewed every four years, is £2,763. Thus, in the "best-case scenario", the lower limit of the incremental cost of cochlear implantation is estimated to be £40,951.

Table 10 lists the lifetime utility gains predicted by each HRQL instrument with future benefits discounted at 6% *per annum*. The EQ-5D estimates a gain of 0.44 QALYs, compared to 2.85 QALYs with the HUI-3. In the best-case scenario, life-time cost-utility ratios are estimated to range from £92,982/QALY for the EQ-5D to £14,368/QALY for the HUI-3.

Discussion

Cost-utility analysis provides a basis for prioritising interventions so as to maximise the utility of a population. Commissioners of health care may tacitly adopt a criterion value of cost-utility beyond which an intervention would be considered competitive enough to warrant consideration for funding. The figure of £20,000 per QALY is sometimes mentioned. The importance of justifying the choice of instrument for measuring the gain in utility that accompanies CI is summarised in Figure 3. Depending on the choice, the resulting ratios either exceed the criterion, straddle it, or fall below it.

Ideally, in selecting an instrument for measuring the change in utility accompanying an intervention, one would choose an instrument that displayed good construct and convergent validity together with high responsiveness. The results of statistical tests of these criteria suggest that one would choose one of the versions of the Health Utilities Index in preference to the EQ-5D or SF-6D instruments for measuring the change in utility that accompanies CI. Some caution would need to attend that choice, however, for the reasons identified in the following critiques of the EQ5D and HUI instruments.

Critique of the EQ5D instrument

The EQ5D instrument is efficient and precise. Patients in the current study found the questions straightforward. Nonetheless, the instrument is relatively unresponsive to profound hearing impairment and cochlear implantation. This result occurs because the five dimensions are relevant to "health-related" effects of conditions and interventions, whereas profoundly hearing impaired people do not perceive themselves to be in ill health. The point is illustrated by responses on the EQ-VAS which asks about "your own health". Average scores were 76.7 (95% c.i. 73.8 to 79.5) pre-operatively and 77.9 post-operatively (95% c.i. 75.3 to 80.5). The values do not differ significantly from each other. Both are close to the population norms for the age bands from 50-69 (Figure 1 in Kind *et al.,* 1998) which embrace the majority of the patients. Similarly, in completing the EuroQoL questionnaire, many patients declared no problem on any of the five dimensions (35% pre-operatively, 48% post-operatively), thereby declaring themselves to be in a state of "full health". A similar ceiling effect was observed by Brazier *et al.* (1993) who reported that over 95% (of the general population) placed themselves as the highest level of the functional dimensions. One consequence of the potential for patients to score at the highest level of the EuroQoL dimensions (and on the composite score) is that the EQ-5D may be unresponsive to improvements in health status.

Further evidence that profoundly hearing-impaired people do not perceive themselves to be in ill-health came from the individual dimensions of the SF36 where patients reported significantly higher scores on the energy and general-health dimensions than the age-weighted norms, both pre- and post-operatively.

Potentially, a second reason for the lack of responsiveness of the EQ-5D questionnaire stems from the wording of the question about "Usual Activities". Patients in the current sample had been profoundly hearing impaired for between 1 and 30 years prior to implantation. Their usual activities reflected the constraints of profound deafness. The fact that hearing impaired patients are forced to adapt, however, does not mean that they do not experience a relaxation in constraint following CI, and perceive the relaxation to be a benefit — frequently cited benefits of CI are being able to undertake more activities outside the home and being less reliant on other people for help in communication. Patients might have responded differently had they been asked whether their current activities are impeded or about their activities prior to becoming profoundly hearing impaired.

Critique of the HUI instrument

The HUI questionnaire is lengthier and linguistically more challenging than the EQ5D. Regrettably, we are not allowed by the originators to reproduce the questions here, the levels of each of the dimensions are however shown in Table 11. To take an extreme example, the six levels on the ambulation dimension, derived from patient responses, include "Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and required a wheelchair to get around the neighbourhood" and "Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood" (Feeny *et al.* 1995). Compare "I have some problems in walking about" in the mobility dimension of the EQ-5D. Complexity of wording may be responsible for the fact that the Emotion Attribute of the HUI was not responsive to CI, in contrast to the Anxiety/Depression dimension of EuroQoL and the Mental Health dimension of SF36, all three of which include questions about anxiety and depression.

There are ambiguities in the wording of the levels of the hearing attribute. It is not clear whether the words "…hear what is said…" (Feeny *et al.* 1995) should be interpreted as "detect the sounds made by the talker" or "understand the words spoken by the talker". The distinction is important, given that patients can often detect sounds which they cannot interpret. The ambiguity, together with the linguistic complexity of the question, may explain why three patients placed themselves at the highest level of the hearing attribute pre-operatively.

The definition of one of the six levels of the Hearing Attribute in HUI-3 describes an anomalous state which rarely, if ever, occurs in practice. Some levels of other attributes may also occur with similarly low prevalence. Although they add to the precision with which a health state can be described, they increase the complexity of the questionnaire and of the task of valuing health states.

Finally, we should mention why the HUI-2 and HUI-3 scoring algorithms result in different estimates of utility (Table 8). There are three reasons for the difference: (i) Different weights are given to the levels of the attributes, and different algorithms are employed to combined the weights into an estimate of utility. (ii) Different questions are used to determine a respondent's level on the pain attribute. (iii) Responses to questions about vision, hearing, and speech are combined into a single Sensation Attribute in HUI-2, but are treated independently in HUI-3. This change eliminates an anomaly in HUI-2 where a patient who reports the lowest level of function on any one of the three attributes is placed at the lowest level of function on the sensation attribute, irrespective of the other two levels. As a result, no change would be registered in the health status of a deaf-blind patient who underwent an intervention that restored sight or hearing. Yet, in principle, there is no reason why someone who is deaf-blind should not benefit from interventions that improve their hearing and speech.

Conclusion

None of the instruments considered here is ideal for measuring the gain in utility that accompanies cochlear implantation for profoundly post-lingually deafened adults. Statistical tests of convergent validity suggest that the HUI-3 may be sensitive to more of aspects of cochlear implantation that improve a person's quality of life than the EQ-5D or SF-6D. Consequently, the HUI-3 might be adopted as the primary outcome measure, despite its deficiencies and its relative unfamiliarity to audiences in the UK. Groups studying the cost-utility of CI in North America (Palmer *et al.* 1999 and Cheng *et al.* 1999) have also employed the HUI, but have obtained supplementary estimates of the gain in utility (Cheng *et al.* 2000). The present results show that the HUI-3 is responsive to the primary purpose, and main benefit, of CI, which is a large effect, while the two alternative instruments respond largely to the secondary benefit, which is a small effect. In adopting the HUI-3, however, it may be appropriate to set limits on the plausibility of estimates of cost-utility that are wider than conventional statistical confidence intervals.

Turning to the broader issue broached in the introduction to this paper, it is a concern that the two measures of health status that are most widely used in the UK are not responsive to the main consequence of profound hearing impairment — difficulty perceiving speech. The high prevalence of hearing disorders in the adult population, coupled with the commitment of the Department of Health to “modernise hearing-aid services” (NHS Executive, 2000), could justify attempts to achieve greater precision in measuring the changes in quality of life that accompany sensory disorders in general, and hearing impairment in particular.

References

Brazier JE, Harper R, Thomas K, Jones N, Underwood T (1998) Deriving a preference based single index measure from the SF-36 Journal Clinical Epidemiology 51 (11):1115-1129

Brazier J. Jones N. Kind P. (1993) Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Quality of Life Research. 2: 169-180.

Brazier, J, Deverill M, Green C, Harper R, Booth A. (1999) A review of the use of health status measures in economic evaluation. Health Technology Assessment.Vol. 3: No.9.

Brazier J. Roberts J. Deverill M. (submitted) The Estimation of a Preference-Based Measure of Health from the SF-36. Paper submitted for publication.

Brooks R, with the EuroQol group (1996). EuroQol: the current state of play. Health Policy. 37: 53-72.

Cheng AK, Niparko JK. (1999) Cost-utility of the cochlear implant in adults: a meta-analysis. Arch Otolaryngol Head Neck Surg. 125: 1214-8.

Cheng A. Rubin H. Powe N. Mellon N. Francis H. Niparko J. (2000) Cost Utility Analysis of the Cochlear Implant in Children. JAMA. 284 (7): 850-856.

Coles, R (1997) Tinnitus. In D Stephens (Ed.) Scott-Brown's Otolaryngology, Part 6: Adult Audiolgy. Oxford: Butterworth-Heinemann. pp. 2/18/1-2/18/34.

Davis, AC (1995) Hearing in Adults. London: Whurr.

Dolan P. Gudex C. Kind P. Williams A. (1995) A Social Tariff for the EuroQol: Results from a UK General population survey. CHE Discussion Paper 138.

Dolan P. Gudex C. Kind P. Williams A. (1996a) The time trade off method: results from a general population study. Health Economics 1996; 5: 141-154.

Dolan P. Gudex C. Kind P. Williams A. (1996b) Valuing health states: A comparison of methods. Journal of Health Economics 1996; 15: 209-231.

Dolan P. Sutton M. (1997) Mapping Visual Analogue Scale health state valuations onto standard gamble and time trade off values. Social Science and Medicine. 44: 1519 – 1530.

Feeney DH, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. (1992) A comprehensive multi-attribute system for classifying the health status of survivors from childhood cancer. J Clin Oncol. 10: 923-928

Feeny D. Furlong W. Boyle M. Torrance G. (1995) Multi- Attribute Health status Classification systems: Health Utilities Index. 7 (6): 490-502.

Furlong W. Feeny D. Torrance G. Goldsmith C. DePauw S. Zhu Z. Denton M. Boyle M (1998) Multiplicative Multi-attribute Utility Function for the Health Utilities Index Mark 3 (HUI 3) system: A Technical Report, McMaster University for Health Economics and Policy Analysis Working Paper 98-11, December 1998.

Gudex C. (1994) Time trade-off user manual. Props and self completion method. Centre for Health Economics Occasional Paper Series.

Gudex C. Dolan P. (1996) Valuing Health States: The Effect of Duration. Centre for Health Economics, Discussion Paper 143.

Harris JP, Anderson JP, Novak R. (1995) An outcomes study of cochlear implants in deaf patients. Audiologic, economic, and quality-of-life changes. Arch Otolaryngol Head Neck Surg. 121: 398-404.

Jenkinson C. Layte R. Wright L. Coulter A. (1996) The UK SF-36: An analysis and interpretation manual. Health Services Research Unit. Oxford.

Kind P (1996). The EuroQol Instrument: An Index of Health Related Quality of Life. . Chapter 22, Quality of Life and Pharmacoeconomics. Edited by Spilker B.

Knutson JF, Schartz HA, Gantz BJ, Tyler RS, Hinriches JV, Woodworth GW. Psychological change following 18 months of cochlear implant use. Annals of Otology, Rhinology, and Laryngology 1991; 100: 877-882.

Loomes G. Sugden R. (1982) Regret Theory: An alternative theory of rational choice under uncertainty.. Economic Journal 92.

NHS Executive (2000) Launch of pilot project to provide patients with digital hearing aids. Press release 2000/0029, 18th January 2000. Accessed at http://www.doh.gov.uk/hearingaidproject/press0029.htm on 1st November 2000.

Palmer CS, Niparko JK, Wyatt JR, Rothman M, de Lissovoy G. (1999) A prospective study of the cost-utility of the multichannel cochlear implant. Arch Otolaryngol Head Neck Surg. 125: 1221-8.

Robinson K. Gatehouse S. Browning G. (1996) Measuring patient benefit from otorhinolaryngological surgery and therapy. Annals of Otology, Rhinology and Laryngology. 105 (6):

Sissons (1996) “Tinnitus questionnaire items for measures of severity, maskability, and intrusion/annoyance.” Dissertation presented to the Department of Public Health and Epidemiology, University of Nottingham 1996.

Summerfield AQ, Marshall DH. *Cochlear Implantation in the UK 1990-1994.* London: HMSO Books, 1995.

Summerfield AQ, Marshall DH. Non-use of cochlear implants by post-lingually deafened adults. *Cochlear Implants International* 2000; 1: 18-38.

Taylor R, Paisley S. The clinical and cost effectiveness of advances in hearing aid technology. Report to the National Institute for Clinical Excellence. 2000.

Torrance G. Furlong W. Feeny D. Boyle M. (1995) Multi-attribute Preference Functions. Pharmacoeconomics. 7(6): 503-518.

Torrance G. Feeny D. Furlong W. Barr R. Zhang Y. Wang Q. (1996) Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2. Medical Care. 7: 702-722.

Ware J. (1993) SF-36 Health Survey: Manual and Interpretation guide. Boston. New England Medical Centre.

Williams A. (1995) The Measurement and Valuation of Health: A chronicle. CHE Discussion Paper 134.

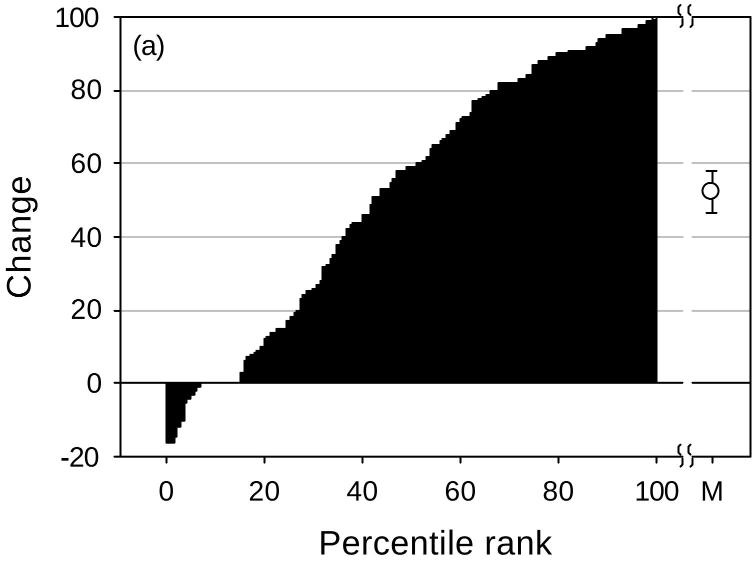
Figure captions

Figure 1. Profile of individual effects of cochlear implantation on (a) speech perception and (b) annoyance due to tinnitus. Patients have been ranked, separately in each plot, in increasing value of their change in score. Size of change has been plotted against each patient's percentile rank. The isolated data point on the right of each graph plots the mean change and its 95% confidence interval.

Figure 2. Profiles of individual effects of cochlear implantation on composite scores derived from HUI-2 (panel (a)), HUI-3 (b), EQ5D-Q (c), EQ5D-VAS (d), SF36-VAS (e), SF36-SG (f), GHSI (g), and GBI (h). Each graph was created in the same fashion as the graphs in Figure 1.

Figure 3. Cost effectiveness plane showing the relationship between incremental cost and incremental effect (QALYs gained) according HRQL instrument. A hypothetical cut off of £20,000 has been used to illustrate which measures suggest cochlear implantation to be cost effective.

Change in speech perception



Change in tinnitus annoyance

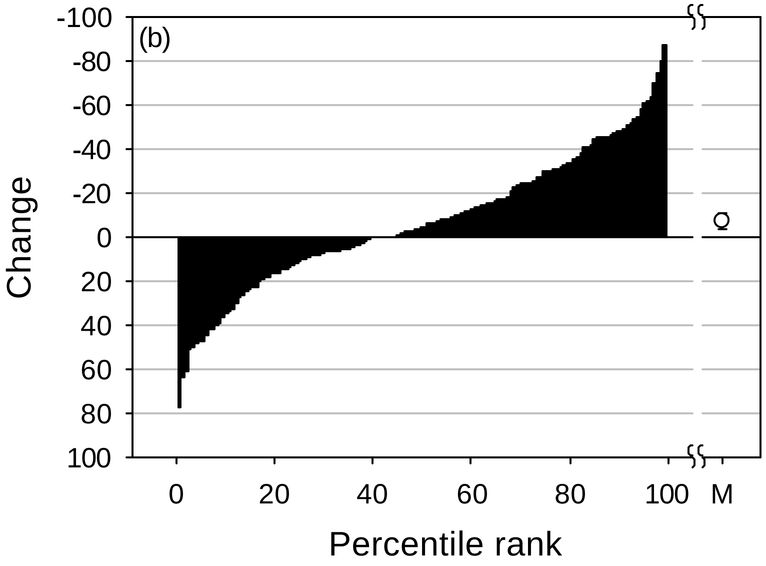


Figure 1

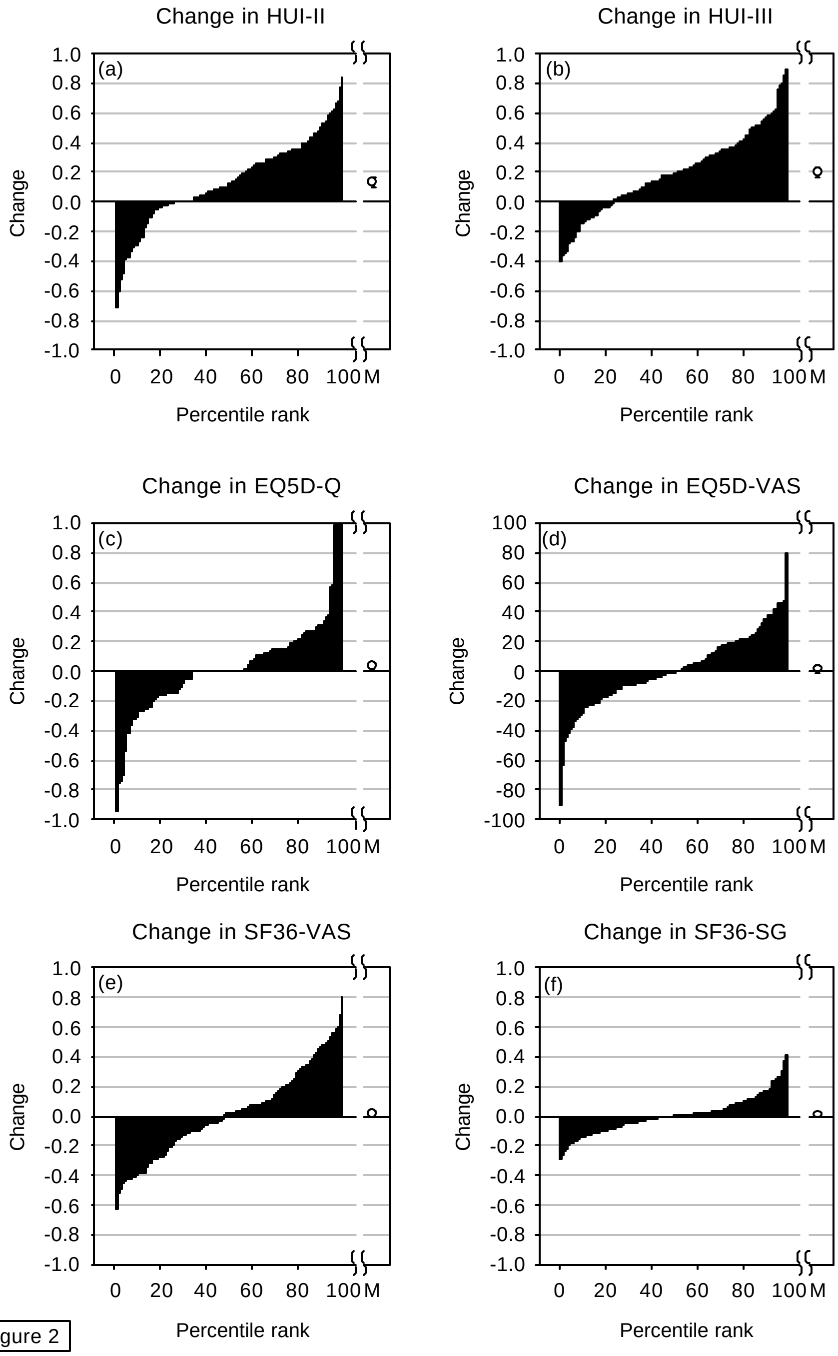
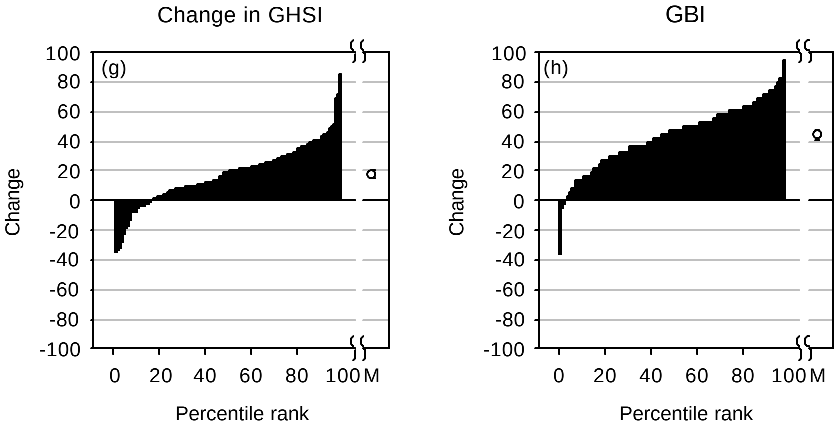

Change in HUI-II
(a)
Change in HUI-III
(b)
Change in EQ5D-Q
(c)
Change in EQ5D-VAS
(d)
Change in SF36-VAS
(e)
Change in SF36-SG
(f)
Change
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
100
80
60
40
20
0
-20
-40
-60
-80
-100
0
20
40
60
80
100
M
Percentile rank

Figure 2



Change in GHSI
GBI
(g)
(h)
100
80
60
40
20
0
-20
-40
-60
-80
-100
Change
0
20
40
60
80
100
M
Percentile rank

Figure 2

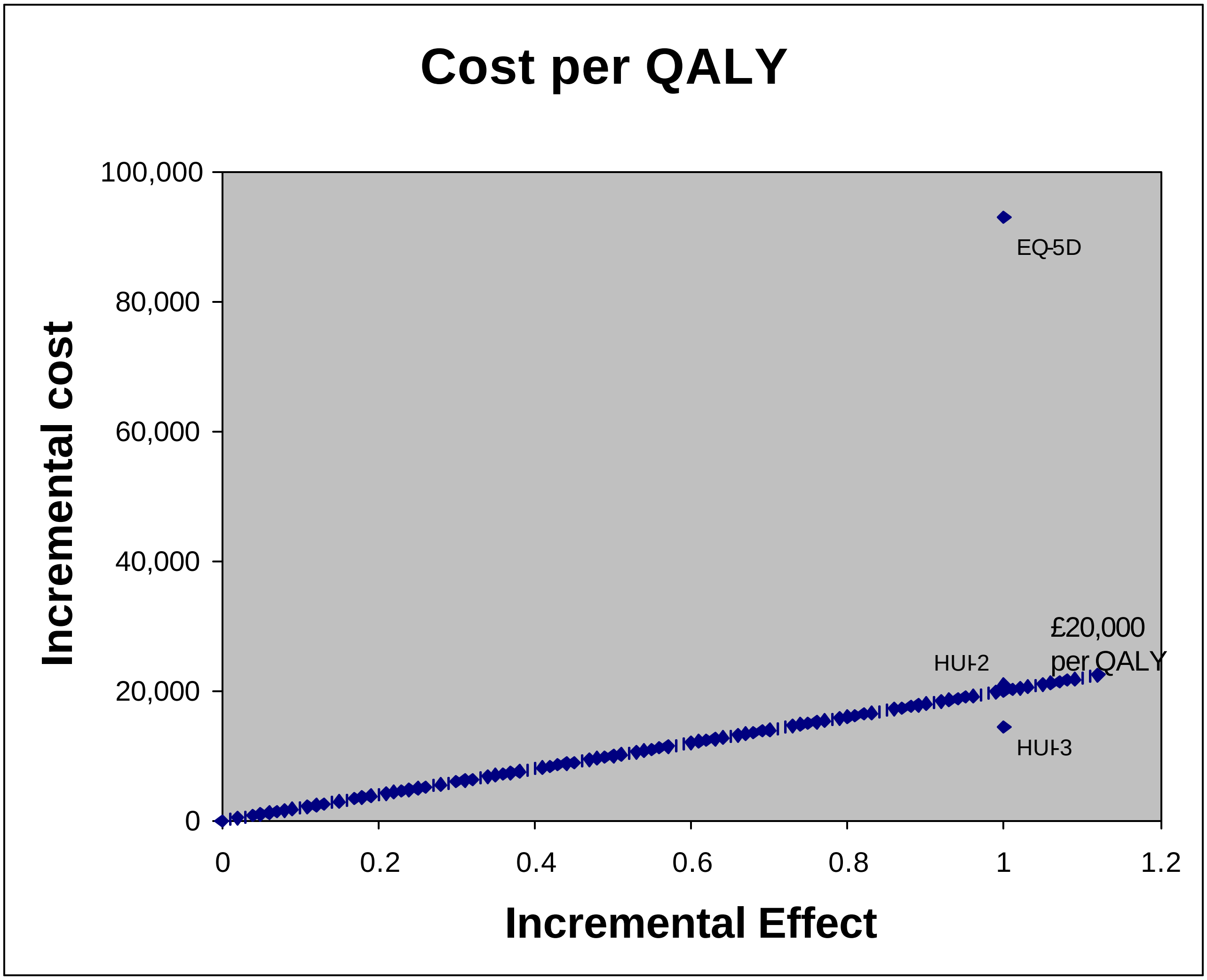

Cost per QALY
Incremental cost
100,000
80,000
60,000
40,000
20,000
0
0
0.2
0.4
0.6
0.8
1
1.2
Incremental Effect
EQ-5D
HUI-2
HUI-3
£20,000 per QALY

Figure 3

Table 1 Age profile of patients at the time of implantation.

Age band (years)	Number of cases
18-19	3
20-29	15
30-39	17
40-49	32
50-59	35
60-69	32
70-79	12
18-79	147

Table 2 Expected numbers of cases (Expected N) reporting a problem in each EuroQoL dimension from age-weighted population norms (Kind *et al.*, 1998, Table 2) and observed numbers of cases (with 95% confidence limits calculated using binomial probabilities) before (Observed N Pre-op) and after (Observed N Post-op) cochlear implantation. Cross-tabulation of observed pre- and post-operative numbers of cases (χ^2) . A tick confirms significance at the p<0.05 level.

	EuroQoL Dimension				
	Mobility	Self care	Usual activities	Pain/Discomfort	Anxiety/Depression
Expected N	28	6	25	52	33
Observed N Pre-op	37	7	34	48	64
(95% c.i.)	(27 to 48)	(3 to 14)	(24 to 45)	(37 to 60)	(52 to 76)
Significant difference from expected value?					✓
Observed N Post-op	39	8	29	58	34
(95% c .i.)	(29 to 51)	(3 to 15)	(20 to 40)	(46 to 70)	(24 to 45)
Significant difference from expected value?	✓				
Change, pre- to post-op	χ^2=0.07	χ^2=0.07	χ^2=0.51	χ^2=1.48	χ^2=13.78
p<0.05?					✓

Table 3 Expected values of each SF36 dimension from age-adjusted population norms (Jenkinson *et al.*, 1996, Table 6) (Expected), values observed pre-operatively (Observed Pre-op) (with 95% confidence interval) and post-operatively (Observed Post-op) (with 95% confidence interval). Change in value on each SF36 dimension between pre- and post-operative states (Change) (with 95% confidence interval). A tick confirms significance at the $p<0.05$ level.

	SF36 Dimension							
	Physical Functioning	Role Physical	Role Mental	Social Functioning	Mental Health	Energy	Pain	Health Perception
Expected	84.2	82.7	83.7	88.2	74.8	61.3	79.7	71.5
Observed Pre-op	79.2	77.2	80.7	80.7	72.8	64.9	79.7	75.6
(95% c.i.)	75.0 to 83.4	71.4 to 83.1	75.6 to 85.8	76.5 to 84.9	69.9 to 75.7	61.6 to 68.2	75.5 to 84.0	72.3 to 78.8
Significant difference from expected value?	✓			✓		✓		✓
Observed Post-op	76.9	75.2	79.4	84.2	77.7	66.9	77.5	75.0
(95% c.i.)	72.4 to 81.4	69.1 to 81.2	73.9 to 84.9	80.6 to 87.8	74.7 to 80.7	63.7 to 70.2	73.3 to 81.7	71.5 to 78.5
Significant difference from expected value?	✓	✓		✓		✓		✓
Change, pre- to post-op.	-2.35	-2.04	-1.36	3.48	4.87	2.01	-2.27	-0.56
(95% c.i.)	-5.44 to 0.75	-8.22 to 4.14	-8.00 to 5.28	-0.18 to 7.14	2.30 to 7.44	-0.78 to 4.80	-5.61 to 1.08	-3.11 to 1.98
Significant change?					✓			

Table 4 Observed numbers of cases who did not place themselves at the top level on each attribute of the HUI-2 (the top 2 levels for the Sensation Attribute) before (Observed N Pre-op) and after (Observed N Post-op) cochlear implantation. Cross-tabulation of observed pre- and post-operative numbers of cases (χ^2) . A tick confirms significance at the p<0.05 level.

	HUI-2 Attribute					
	Sensation	Mobility	Emotion	Cognition	Self-care	Pain
Observed N Pre-op	138	22	52	44	1	23
Observed N Post-op	67	20	42	37	4	37
Change, pre- to post-op	$\chi^2=81.23$	$\chi^2=0.11$	$\chi^2=1.56$	$\chi^2=0.84$	$\chi^2=0$	$\chi^2=5.50$
p<0.05?	✓					✓

Table 5 Observed numbers of cases who did not place themselves at the top level on each attribute of the HUI-3 (the top 3 levels for the Hearing Attribute) before (Observed N Pre-op) and after (Observed N Post-op) cochlear implantation. Cross-tabulation of observed pre- and post-operative numbers of cases (χ^2) . A tick confirms significance at the p<0.05 level.

	HUI3 Attribute							
	Vision	Hearing	Speech	Ambulation	Dexterity	Emotion	Cognition	Pain
Observed N Pre-op	99	132	41	22	2	50	44	62
Observed N Post-op	106	38	33	20	0	53	37	75
Change, pre- to post-op	χ^2=0.79	χ^2=123.24	χ^2=1.16	χ^2=0.11	χ^2=0	χ^2=0.13	χ^2=0.84	χ^2=2.31
p<0.05?		✓						

Table 6 Single-attribute utility scores for each attribute of the Mark 2 Health Utilities Index prior to cochlear implantation (Pre-op) and after cochlear implantation (Post-op), and changes in single-attribute utility values (post-op minus pre-op) (with 95% confidence limits). A tick confirms significance at the p<0.05 level.

	HUI-2 Attribute					
	Sensation	Mobility	Emotion	Cognition	Self-care	Pain
Observed Pre-op	0.34	0.96	0.93	0.96	1.00	0.91
Observed Post-op	0.75	0.96	0.95	0.96	0.98	0.86
Change, pre- to post-op	0.41	0.00	0.02	0.00	0.02	0.05
Upper 95% c.l.	0.47	0.03	0.05	0.02	0.00	0.01
Lower 95% c.l.	0.35	-0.03	-0.01	-0.01	-0.03	-0.11
p<0.05?	✓					

Table 7 Single-attribute utility scores from a population sample (from Grootendorst, Feeny, and Furlong, 2000, Table 1) for each attribute of the Mark 3 Health Utilities Index (Expected), Observed values (with 95% confidence limits) prior to cochlear implantation (Pre-op) and after cochlear implantation (Post-op), and changes in single-attribute utility values (post-op minus pre-op) (with 95% confidence limits). A tick confirms significance at the $p<0.05$ level.

	HUI-3 Attribute							
	Vision	Hearing	Speech	Ambulation	Dexterity	Emotion	Cognition	Pain
Expected	0.974	0.989	0.998	0.988	0.996	0.981	0.959	0.951
Observed Pre-op	0.959	0.211	0.911	0.956	0.993	0.935	0.939	0.899
Upper 95% c.l.	0.969	0.251	0.940	0.975	1.000	0.960	0.960	0.929
Lower 95% c.l.	0.948	0.170	0.882	0.937	0.982	0.910	0.919	0.869
Significant difference from expected value?	✓	✓	✓	✓		✓		✓
Observed Post-op	0.958	0.598	0.948	0.955	1.000	0.953	0.942	0.895
Upper 95% c.l.	0.968	0.631	0.965	0.977	1.000	0.965	0.965	0.921
Lower 95% c.l.	0.948	0.565	0.932	0.933	1.000	0.940	0.920	0.869
Significant difference from expected value?	✓	✓	✓	✓		✓		✓
Change, pre- to post-op.	-0.001	0.388	0.037	-0.001	0.007	0.018	0.003	-0.004
Upper 95% c.l.	0.007	0.433	0.068	0.019	0.018	0.043	0.021	0.019
Lower 95% c.l.	-0.009	0.342	0.006	-0.021	-0.003	-0.007	-0.014	-0.026
Significant change?		✓	✓					

Table 8 For each of 10 measures, mean change (post-implantation - pre-implantation), standard deviation (SD), standard error of the mean (SE), lower and upper 95% confidence limits, significance at p<0.05, and effect size (absolute value of mean divided by SD).

	Speech reception	Tinnitus	HUI-II	HUI-III	EQ5D-Q	EQ5D-VAS	SF 6D-VAS	SF 6D-SG	GHSI	GBI
Mean difference	52.274	-7.128	0.135	0.197	0.030	1.261	0.014	0.006	17.475	43.950
SD	34.518	21.332	0.201	0.197	0.187	15.337	0.155	0.065	15.159	21.523
SE	2.847	1.759	0.017	0.016	0.015	1.265	0.013	0.005	1.255	1.781
Upper 95% c.i.	57.854	-3.679	0.168	0.229	0.061	3.741	0.039	0.016	19.934	47.441
Lower 95% c.i.	46.694	-10.576	0.103	0.165	0.000	-1.218	-0.011	-0.005	15.016	40.458
p<0.05	✓	✓	✓	✓	✓				✓	✓
Effect size	1.514	0.334	0.673	0.998	0.163	0.082	0.092	0.089	1.153	2.042
N	147	147	147	147	147	147	147	147	147	147

Table 9 Kendal rank correlation coefficients (tau), probabilities of occurrence (p), and significance levels (** p<0.01, *** p<0.001), computed between each of 10 measures and measures of speech reception and annoyance due to tinnitus before implantation and after implantation, and corresponding correlations between changes in measures. The GBI is a direct measure of change in health-related quality of life.

		Measure									
		Speech	Tinnitus	HUI-3	HUI-2	EQ5D-Q	EQ5D-VAS	SF 6D-VAS	SF 6D-SG	GHSI	GBI
Before implantation											
Speech Perception	tau		-0.053	0.137	0.178	0.016	0.053	0.106	0.102	-0.108	
	p		0.412	0.033	0.007 **	0.816	0.402	0.096	0.114	0.097	
Tinnitus	tau	-0.053		-0.242	-0.230	-0.226	-0.225	-0.229	-0.218	-0.118	
	p	0.412		0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.040	
After implantation											
Speech Perception	tau		0.002	0.189	0.195	0.127	0.044	0.150	0.151	0.167	
	p		0.978	0.001 **	0.001 **	0.038	0.438	0.008 **	0.008 **	0.003 **	
Tinnitus	tau	0.002		-0.206	-0.201	-0.246	-0.236	-0.214	-0.209	-0.254	
	p	0.978		0.000 ***	0.001 **	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	
Change in score											
Speech Perception	tau		-0.056	0.191	0.184	0.121	-0.004	0.077	0.087	0.203	0.308
	p		0.327	0.001 **	0.001 **	0.042	0.944	0.170	0.124	0.000 ***	0.000 ***
Tinnitus	tau	-0.056		-0.069	-0.111	-0.189	-0.179	-0.124	-0.119	-0.129	-0.102
	p	0.327		0.218	0.051	0.002 **	0.001 **	0.028	0.036	0.024	0.075

Table 10 Discounted costs of providing CI projected over thirty years, QALYs gained calculated from the gain in utility estimated by three health-status instruments, and cost-utility ratios.

Gain in utility	Measure	Life Expectancy 30 years
	Incremental Cost (discounted 6%)	£40,951
+0.03 (EuroQoL)	QALY gain (discounted 6%)	0.44
	C-U ratio	£92,982
+0.14 (HUI-2)	QALY gain (discounted 6%)	1.96
	C-U ratio	£20,912
+0.20 (HUI-3)	QALY gain (discounted 6%)	2.85
	C-U ratio	£14,368
+0.03 (SF-6D SG)	QALY gain (discounted 6%)	0.084
	C-U ratio	£486,930

Table 11: Health Utilities Index Mark 3: Health status classification system (Feeny *et al.* 1995)

VISION

Level	Description of Level
1	Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses
2	Able to see well enough to read ordinary newsprint and and recognize a friend on the other side of the street, but with glasses
3	Able to see well enough to read ordinary newsprint with or without glasses but unable to recognize a friend on the other side of the street, even with glasses or contact lenses
4	Able to recognize a friend on the other side of the street with or without glasses but unable to read ordinary newsprint, even with glasses or contact lenses
5	Unable read ordinary newsprint and unable to recognize a friend on the other side of the street, even with glasses
6	Unable to see at all

HEARING

Level	Description of Level
1	Able to hear what is said in a group conversation with atleast three other people, without a hearing aid
2	Able to hear what is said in a conversation with 1 other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least 3 other people
3	Able to hear what is said in a conversation with 1 other person in a quiet room with a hearing aid, and able to hear what is said in a group conversation with at least 3 other people with a hearing aid
4	Able to hear what is said in a conversation with 1 other person in a quiet room without a hearing aid, but unable to hear what is said in a group conversation with at least 3 other people even with a hearing aid
5	Able to hear what is said in a conversation with 1 other person in a quiet room with a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid
6	Unable to hear at all

SPEECH

Level	Description of Level
1	Able to be understood completely when speaking with strangers or people who know me well
2	Able to be understood partially when speaking with strangers but able to be understood completely when speaking with people who know me well
3	Able to be understood partially when speaking with strangers or people who know me well
4	Unable to be understood when speaking with strangers but able to be understood but able to be understood partially by people who know me well.
5	Unable to be understood partially when speaking to other people (or unable to speak at all)

AMBULATION

Level	Description of Level
1	Able to walk around the neighbourhood without difficulty, and without walking equipment.
2	Able to walk around the neighbourhood with difficulty, but does not require walking equipment or the help of another person
3	Able to walk around the neighbourhood with walking equipment, but without the help of another person
4	Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood
5	Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood
6	Cannot walk at all.

DEXTERITY

Level	Description of Level
1	Full use of 2 hands and 10 fingers
2	Limitations in the use of hands or fingers, but does not require special tools or help of another person
3	Limitations in the use of hands or fingers, is independent with the use of special tools and does not require the help of another person
4	Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with the use of special tools)
5	Limitations in use of hands or fingers, requires the help of another person for most tasks (not independent even with the use of special tools).
6	Limitations in use of hands or fingers, requires the help of another person for all tasks (not independent even with the use of special tools).

EMOTION

Level	Description of Level
1	Happy and interested in life
2	Somewhat happy
3	Somewhat unhappy
4	Very unhappy
5	So unhappy that life is not worthwhile

COGNITION

Level	Description of Level
1	Able to remember most things, think clearly and solve everyday problems
2	Able to remember most things, but has a little difficulty when trying to think and solve day to day problems.
3	Somewhat forgetful, but able to think clearly and solve day to day problems
4	Somewhat forgetful, and has a little difficulty when trying to think or solve day to day problems
5	Very forgetful, and has great difficulty when trying to think or solve day to day problems
6	Unable to remember anything at all, and unable to think or solve day to day problems

PAIN and DISCOMFORT

Level	Description of Level
1	Free from pain and discomfort
2	Mild to moderate pain that prevents no activities
3	Moderate pain that prevents a few activities
4	Moderate to severe pain that prevents some activities
5	Severe pain that prevents most activities