

Choosing between measures: comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints

Janneke P. C. Grutters · Manuela A. Joore ·
Frans van der Horst · Hans Verschuure ·
Wouter A. Dreschler · Lucien J. C. Anteunis

Accepted: 26 June 2007 / Published online: 24 July 2007
© Springer Science+Business Media B.V. 2007

Abstract

Objectives To generate insight into the differences between utility measures EuroQol 5D (EQ-5D), Health Utilities Index Mark II (HUI2) and Mark III (HUI3) and

their impact on the incremental cost-effectiveness ratio (ICER) for hearing aid fitting

Methods Persons with hearing complaints completed EQ-5D, HUI2 and HUI3 at baseline and, when applicable, after hearing aid fitting. Practicality, construct validity, agreement, responsiveness and impact on the ICER were examined.

Results All measures had high completion rates. HUI3 was capable of discriminating between clinically distinctive groups. Utility scores ($n = 315$) for EQ-5D UK and Dutch tariff (0.83; 0.86), HUI2 (0.77) and HUI3 (0.61) were significantly different, agreement was low to moderate. Change after hearing aid fitting ($n = 70$) for HUI2 (0.07) and HUI3 (0.12) was statistically significant, unlike the EQ-5D UK (0.01) and Dutch (0.00) tariff. ICERs varied from €647,209/QALY for the EQ-5D Dutch tariff to €15,811/QALY for HUI3.

Conclusion Utility scores, utility gain and ICERs heavily depend on the measure that is used to elicit them. This study indicates HUI3 as the instrument of first choice when measuring utility in a population with hearing complaints, but emphasizes the importance of a clear notion of what constitutes utility with regard to economic analyses.

J. P. C. Grutters · M. A. Joore
Department of Clinical Epidemiology and Medical Technology
Assessment, University Hospital Maastricht, PO Box 5800, 6202
AZ Maastricht, The Netherlands

M. A. Joore
e-mail: mjo@kemta.azm.nl

J. P. C. Grutters · L. J. C. Anteunis
Department of Otorhinolaryngology, Head and Neck Surgery,
University Hospital Maastricht, PO Box 5800, 6202 AZ
Maastricht, The Netherlands

L. J. C. Anteunis
e-mail: l.anteunis@skno.azm.nl

Present Address:

J. P. C. Grutters (✉)
Maastricht Clinic, PO Box 5800, 6202 AZ Maastricht,
The Netherlands
e-mail: janneke.grutters@maastro.nl

F. van der Horst
Department of General Practice, Maastricht University,
PO Box 616, 6200 MD Maastricht, The Netherlands
e-mail: frans.vanderhorst@hag.unimaas.nl

H. Verschuure
Audiological Center, Erasmus Medical Center Rotterdam,
PO Box 2040, 3000 CA Rotterdam, The Netherlands
e-mail: j.verschuure@erasmusmc.nl

W. A. Dreschler
Department of Clinical and Experimental Audiology,
Academic Medical Center Amsterdam, PO Box 22660,
1100 DD Amsterdam, The Netherlands
e-mail: w.a.dreschler@amc.nl

Keywords Hearing loss · Costs and cost analysis ·
Quality-adjusted life years · Questionnaires · Outcome
assessment (health care)

Abbreviations

BEPTA	Better ear pure tone average hearing loss
CI	Confidence Interval
ENT	Ear Nose and Throat
EQ-5D UK tariff	EuroQol 5d with British utility scoring function

EQ-5D Dutch tariff	EuroQol 5d with Dutch utility scoring function
GP	General Practitioner
HRQoL	Health related quality of life
HUI2	Health Utilities Index Mark II
HUI3	Health Utilities Index Mark III
ICC	Intra-class correlation
ICER	Incremental cost-effectiveness ratio
IR	Interquartile Range
QALY	Quality adjusted life year
SG	Standard Gamble
TTO	Time trade-off
VAS	Visual analogue scale

scores derived with the EQ-5D UK tariff [26], the EQ-5D Dutch tariff [27], the Health Utilities Index Mark 2 (HUI2) and HUI3 [28] are compared to generate further insight into the differences between the measures, and the impact on the ICER for hearing aid fitting. In particular, this article considers:

- (1) practicality of using the EuroQol and HUI in a population with hearing complaints;
- (2) construct validity of the EQ-5D UK tariff, EQ-5D Dutch tariff, HUI2 and HUI3;
- (3) agreement between the EQ-5D UK tariff, EQ-5D Dutch tariff, HUI2 and HUI3;
- (4) responsiveness of the EQ-5D UK tariff, EQ-5D Dutch tariff, HUI2 and HUI3 after hearing aid fitting;
- (5) and the impact of on the ICER for hearing aid fitting.

Introduction

Hearing loss affects the ability to exchange information and therefore affects a person's quality of life [1]. Of the western adult population, approximately 15% are hearing impaired [2], and for these persons hearing aid use has proven to be effective [3, 4]. The growth of the elderly population has far-reaching implications for auditory health service delivery and expenditure since the prevalence increases heavily with age. As a result, increasingly more economic evaluations are undertaken on interventions such as hearing aid fitting. In economic evaluations health related quality of life (HRQoL) should be measured with a preference-based utility measure [5]. Until recently, the benefits of hearing aid fitting had not translated into a statistically significant improvement in health state utility [6–8]. In 2004 Barton et al [9] confirmed this for the EuroQol 5D (EQ-5D), but did find statistically significant utility gain after hearing aid fitting with the Health Utilities Index Mark III (HUI3). This difference may result from differences in the descriptive system, and the way the scoring function is derived.

Also with regard to other conditions, a number of studies have found that different utility measures tend to lead to different utility scores [10–22]. In general there is need for head-to-head comparisons of utility measures, in order to assess the implications for the interpretation and comparability of economic analyses, especially in conditions where only subtle changes after treatment are expected [23, 24]. In these comparisons the comparison of different tariffs should also be incorporated. But most importantly, the impact of the differences between measures on incremental cost-effectiveness ratios (ICERs) should be made apparent. Surprisingly, the latter has not yet received much attention.

The objective of this article is to compare the two most frequently used utility measures in economic analyses [25] in a Dutch population with hearing complaints. Utility

Methods

The EQ-5D, HUI2 and HUI3

The five questions of the EQ-5D descriptive system each represent one dimension of health-related quality of life (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) [29]. Each question has three levels and the questions together classify persons into one of 243 health states. The commonly used scoring function is based on a British study (EQ-5D UK tariff) [26], with preferences derived with the time trade-off (TTO) method, in a representative sample of the UK population of 2,997 respondents. The scoring function is additive and the possible range of utility scores is –0.59 to 1.00. Recently, Lamers et al [27] developed a Dutch scoring function for the EQ-5D (EQ-5D Dutch tariff), based on TTO in a sample of 298 respondents, with utility scores ranging from –0.33 to 1.00.

The 15 questions of the HUI descriptive system classify respondents into either HUI2 or HUI3 health states. The HUI system focuses on health 'within the skin', meaning that they purely focus on impairment and not on the social context of the impairment [28]. HUI2 consists of seven attributes (sensation, mobility, emotion, cognition, self-care, pain and fertility), with three to five levels, leading to 24,000 possible health states. The optional fertility attribute was not used in the present study. The multiplicative scoring function was derived using standard gamble (SG) and Visual Analogue Scale (VAS) in a random sample of 293 Canadian respondents and utility scores range from –0.03 to 1.00 [30]. HUI3 consists of eight attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain), with five to six levels and leads to 972,000 possible health states. The multiplicative scoring function was derived from SG and VAS, in a random sample of the

Canadian general population ($n = 504$), resulting in possible utility scores varying from -0.36 to 1.00 [31].

Study population and data collection

Data were collected as part of a before-after study examining direct hearing aid provision versus provision by referral [32]. The study was carried out in 2004–2005 in three regions (Maastricht, Rotterdam and Amsterdam) in the Netherlands. Persons with hearing complaints were recruited from the participating Ear Nose and Throat (ENT) departments, audiological centres and hearing aid dispensers. Both EQ-5D and HUI were administered at the first visit to the ENT department. Questionnaires were completed at the department, and respondents were helped by a trained interviewer if requested. Patients who were fitted with a hearing aid were asked to attend the ENT department for a follow-up visit to evaluate the hearing aid fitting. During the evaluation at the ENT department this subset of patients completed both questionnaires for a second time. Again, questionnaires were completed at the department, and respondents were helped by a trained interviewer if requested.

Practicality of the questionnaires

Especially in an elderly population, an important aspect of a utility measure is the ease of completion. The practicality of using the EQ-5D and the HUI in a population with hearing complaints was therefore assessed by the completion rate, using a Wilcoxon Signed Ranks Test to test whether the completion rates were significantly different. Additionally we examined the item non-response.

Construct validity

In absence of a gold standard to measure health state utility, there is no clear technique to determine the construct validity of utility measures. A way to examine the construct validity is to examine whether utility scores are different for distinctive groups [33, 34]. Other studies have demonstrated differences of quality of life scores by sex [33, 35, 36] and age [35–38]. Comparisons were made between EQ-5D UK and Dutch tariff, HUI2 and HUI3 scores by age (above versus below median) and sex. It was expected that persons of a higher age and females have lower utility scores. In addition, the respondents were divided into five clinically distinctive groups, based on their hearing loss and hearing aid use. Hearing loss was defined as the better ear pure tone average (BEPTA) hearing loss for the frequencies 1000, 2000 and 4000 Hz. The groups were: persons who were

not entitled to reimbursement of a hearing aid (BEPTA < 35 dB); persons who were entitled to reimbursement (BEPTA ≥ 35 dB) but did not apply for a hearing aid (non-applicants); first time hearing aid applicants; experienced hearing aid users who were about to have a new hearing aid fitted (re-applicants); and experienced hearing aid users who did not have a new hearing aid fitted.

It was expected that persons with a BEPTA smaller than 35 dB would have a higher quality of life score than persons in the other four groups, because they are likely to experience less problems with hearing. It was also expected that non-applicants had a higher utility score than first time applicants, since the latter group is expected to experience more hearing complaints, resulting in fitting a hearing aid as a solution for their hearing problems.

Descriptive summary statistics were provided and normality was tested for all data using the Kolmogorov–Smirnov test. Kruskal–Wallis one way ANOVA and pairwise comparison tests (Mann–Whitney U) were used to explore the differences between the groups.

Agreement

To assess agreement between the measures, a Wilcoxon Signed-Rank test and a Paired t -test were used to test whether the scores of the EQ-5D (UK and Dutch tariff), HUI2 and HUI3 had the same distribution and mean. Correlations (Kendall's Tau) and the Intra-class Correlation Coefficient (ICC) were computed. The ICC was based on a two-way mixed effect model, such that the subject effect was random and the instrument effect was fixed, and computed at the individual patient level. An ICC below 0.75 implies poor to moderate agreement; above 0.75 implies good agreement [34].

Responsiveness

Responsiveness was tested in the subpopulation with a valid score both before and after hearing aid fitting. Effect size and standardized response mean were calculated. Effect size is the change in score divided by the standard deviation of scores at baseline. Standardized response mean is the change in score divided by the standard deviation of the change in score. Both were interpreted using benchmarks for effect size: 0.20 through 0.49 is interpreted as small, 0.50 through 0.79 as moderate and ≥ 0.80 as large [39]. Also, a Wilcoxon Signed Ranks Test and a Paired t -test were conducted on the before and after scores. The change in score after hearing aid fitting was tested for differences between first time hearing aid applicants and re-applicants. It was expected that re-applicants had a

smaller change in utility, since they already had a hearing aid and therefore expectedly less hearing problems at baseline.

Impact on the ICER

To illustrate the impact on the ICER, straightforward one year ICERs of hearing aid fitting versus no hearing aid fitting (doing nothing) were calculated. Quality Adjusted Life Years (QALYs) were calculated using the area under the curve method, for the EQ-5D (UK and Dutch tariff), HUI2 and HUI3 [5]. Baseline utility scores were used to estimate the effects of ‘no hearing aid fitting’. Costs of no hearing aid fitting were assumed to be zero. This was compared to the costs and effects after hearing aid fitting. The costs of hearing aid fitting were calculated prospectively for each respondent, including General Practitioner (GP) visit(s), ENT visit(s) and hearing aid(s). We used standard costs for medical consumption [40], and the actual price of the hearing aid. As costs and utility scores are generally not normally distributed, a non-parametric bootstrap sampling method was used to calculate the 95% confidence interval around the ICERs [41]. Cost-effectiveness acceptability curves for all measures were created to characterize the likelihood that hearing aid fitting will be deemed cost-effective based on the incremental costs and outcomes, for a range of ceiling ratios. Ceiling ratios reflect societies’ maximum willingness to pay for a unit of outcome.

Results

Practicality

Of the 337 persons with hearing complaints included in the study, in total 315 (93%) fully completed both the EQ-5D and the HUI descriptive system at baseline. Each item on the EQ-5D had six missing values, except for mobility ($n = 5$). Regarding the HUI, all questions had six or seven missing values, except for the questions on hearing in a group conversation ($n = 12$), pain and discomfort ($n = 10$), and hearing in a conversation with one other person ($n = 8$). Although completion rates were high for both questionnaires, EQ-5D was fully completed by 328 persons (97%), which is significantly more than the 318 persons (94%) who completed the HUI (P -value 0.012).

Of the 315 persons who completed both EQ-5D and HUI at baseline, 173 persons (55%) had a hearing aid fitted. Of them, 82 (47%) attended the ENT department for the follow-up visit after hearing aid fitting. Ninety-one respondents (53%) who had a hearing aid fitted did not show at the follow-up visit because they had not finished

their hearing aid fitting before the end of the study ($n = 37$), or because they had a hearing aid fitted at a dispenser not participating in the study ($n = 54$), and were therefore lost to follow up. Of the 82 respondents who did attend the follow-up visit, 70 (85%) fully completed both the EQ-5D and HUI descriptive system. Each item on the EQ-5D had five missing values, except for pain/discomfort ($n = 6$). Regarding the HUI, the hearing questions both had four missing values, and the other questions had four to seven missing values. Seventy-six persons (93%) fully completed the EQ-5D, while 71 persons (87%) fully completed the HUI. This difference is not statistically significant (P -value 0.059).

Construct validity

Mean age of the respondents was 69.6 years (sd 8.9; median 70), and BEPTA was on average 42 dB. The respondents were divided into groups below 70 years old ($n = 156$) versus 70 years and older ($n = 159$) and male ($n = 189$) versus female ($n = 126$), see Table 1. The scores on the EQ-5D, HUI2 and HUI3 were not normally distributed (Kolmogorov–Smirnov test, P -value 0.000). Only the EQ-5D detected differences in reported health by both age and sex. HUI3 detected a difference in utility between the age groups, HUI2 did not detect any differences.

Furthermore, the respondents were divided into five clinically distinctive groups: BEPTA < 35 dB ($n = 69$), non-applicants ($n = 46$), first time hearing aid applicants ($n = 108$), re-applicants ($n = 65$), and experienced hearing aid users not applying for a new hearing aid ($n = 12$). The latter group was excluded from the analysis because of the small sample size. Fifteen persons could not be classified into a clinical group because they were lost to follow-up after the first visit.

Based on the EQ-5D and HUI2, no distinction could be made between any of the clinically distinctive groups. A logarithmic transformation was performed on the EQ-5D and HUI2 data to compensate for skewness. Even after transformation, and also when correcting for age, sex and BEPTA, no differences were found between the groups. Only HUI3 scores demonstrated a significant difference between the clinically distinctive groups (Kruskal–Wallis; P -value 0.004). More specifically, HUI3 found significant differences between persons with a BEPTA < 35 dB and first time applicants (Mann–Whitney U; P -value 0.002), and between persons with a BEPTA < 35 dB and re-applicants (P -value 0.001). HUI3 did not confirm our expectation that non-applicants had significantly higher utility scores than first time applicants. As expected, non-applicants stated less problems on the hearing attribute than first time applicants, but they also stated more problems on the ambulation and pain attributes.

Table 1 Comparison of mean scores, median scores and interquartile range (IR) at baseline according to different characteristics of the respondents

	N	Age		Sex	% Male	BEPTA		EQ-5D UK tariff		EQ-5D Dutch tariff		HUI2		HUI3	
		Mean	IR			Mean	IR	Mean	IR	Mean	IR	Mean	IR	Mean	IR
Age															
		<70	156	61.9	64%	37.4	0.86	0.94 (0.24)	0.88	0.95 (0.20)	0.78	0.79 (0.15)	0.64	0.65 (0.36)	
		≥70	159	77.0	57%	47.1	0.81	0.80 (0.27)	0.84	0.84 (0.23)	0.76	0.80 (0.17)	0.58	0.59 (0.38)	
<i>P-value*</i>								0.010		0.025		0.134		0.034	
Sex		Male	189	69.3	100%	43.4	0.85	0.88 (0.26)	0.87	0.90 (0.19)	0.77	0.80 (0.13)	0.61	0.62 (0.37)	
		Female	126	69.9	0%	40.5	0.81	0.80 (0.27)	0.83	0.84 (0.23)	0.76	0.78 (0.18)	0.61	0.63 (0.39)	
<i>P-value*</i>								0.019		0.012		0.575		0.965	
Clinically distinctive groups		BEPTA <35 dB	69	61.9	54%	21.3	0.84	0.81 (0.27)	0.86	0.84 (0.20)	0.80	0.80 (0.20)	0.70	0.66 (0.37)	
		Non-applicants	46	74.2	59%	42.1	0.80	0.87 (0.27)	0.83	0.90 (0.23)	0.77	0.78 (0.15)	0.62	0.60 (0.41)	
		First time applicants	108	71.2	62%	46.9	0.84	0.85 (0.27)	0.86	0.87 (0.23)	0.76	0.78 (0.16)	0.58	0.61 (0.28)	
		Re-applicants	65	69.6	66%	53.6	0.85	0.85 (0.27)	0.88	0.86 (0.20)	0.77	0.79 (0.17)	0.56	0.61 (0.41)	
<i>P-value**</i>								0.996		0.993		0.264		0.004	

* Mann–Whitney *U* test

** Kruskal–Wallis test

Agreement

A summary of the utility scores is presented in Table 2. Mean utility scores for the population with hearing complaints were higher for the EQ-5D Dutch tariff (mean 0.86; standard deviation 0.18) and EQ-5D UK tariff (0.83; sd 0.21) than for the HUI2 (0.77; sd 0.14) and HUI3 (0.61; sd 0.24). The differences in mean scores and distributions are statistically significant. The UK and Dutch tariff of the EQ-5D and HUI2 and HUI3 were found to have good agreement, since they were strongly correlated (0.90; 0.71) and had an ICC of 0.98 and 0.74, respectively. The scores on all other measures showed statistically significant, but low correlations and their agreement was moderate to poor (Table 3).

Ceiling effects were observed in the EQ-5D UK tariff (Figs. 1 and 2), results were similar for EQ-5D Dutch tariff. As measured with the EQ-5D (both UK and Dutch tariff), 44% of the respondents reported perfect health, despite their hearing complaints. Measured with the HUI2 or HUI3, less than 1% of the respondents reported perfect health. For respondents reporting perfect health on the EQ-5D, mean utility scores were 0.83 on the HUI2 (range 0.35–1.00) and 0.71 on the HUI3 (range 0.06–1.00).

Responsiveness

Responsiveness was tested in the subpopulation with a valid score both before and after hearing aid fitting ($n = 70$). The effect size and standardized response mean of the EQ-5D UK (0.05; 0.05) and Dutch tariff (0.03; 0.02) were less than small. HUI2 and HUI3 were more responsive to change after hearing aid fitting, since both had a moderate effect size (0.64; 0.55) and standardized response mean (0.57; 0.66).

Mean change in utility after hearing aid fitting (Table 4) was highest when measured with the HUI3 (mean 0.12; sd 0.18) and HUI2 (0.07; sd 0.13), while almost no change was measured with the EQ-5D (UK tariff 0.01, sd 0.13; Dutch tariff 0.00, sd 0.12). The change in utility measured with HUI2 and HUI3 is statistically significant (Paired *t*-test, *P*-values 0.005 and 0.000). No change was observed in any attribute of the EQ-5D (Fig. 3a). The change in HUI2 utility score after hearing aid fitting occurred in the sensation attribute (Fig. 3b) and in the HUI3 score in the hearing attribute (Fig. 3c). Almost no change was observed in any of the other attributes of the HUI2 and HUI3.

The mean change in utility score after hearing aid fitting, when measured with HUI2 and HUI3, was higher for first-time hearing aid applicants (0.08; 0.13) than for re-applicants (0.06; 0.10). This outcome was in line with our expectations, but is not significantly different.

Table 2 Utility scores in baseline population with hearing complaints ($n = 315$)

Measure	Minimum	Maximum	Median*	Interquartile range	Mean**	Standard deviation
EQ-5D UK tariff	-0.25	1.00	0.85	0.27	0.83	0.21
EQ-5D Dutch tariff	-0.03	1.00	0.86	0.19	0.86	0.18
HUI2	0.23	1.00	0.79	0.15	0.77	0.14
HUI3	-0.07	1.00	0.62	0.38	0.61	0.24

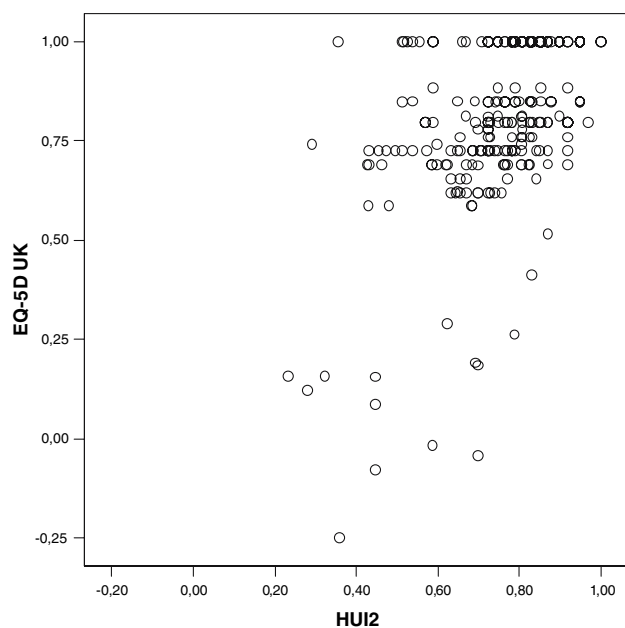
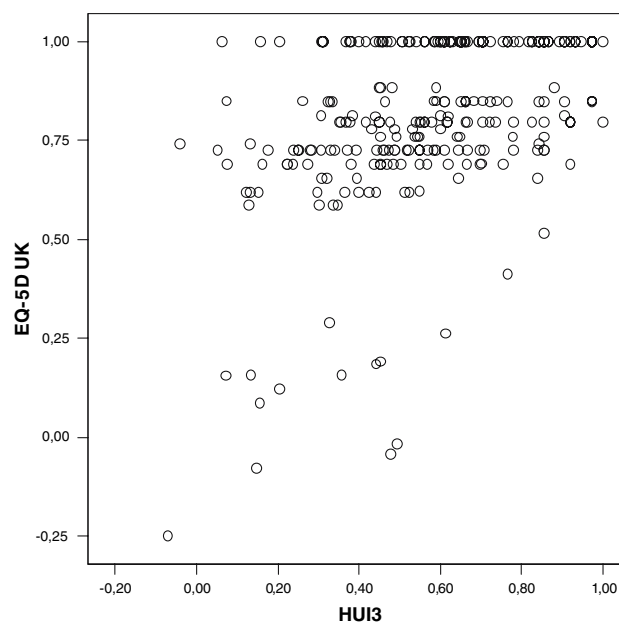
* All statistically significantly different: Wilcoxon Signed-Rank test; $P < 0.01$

** All statistically significantly different: Paired t -test; $P < 0.01$

Table 3 Agreement in the baseline population with hearing complaints ($n = 315$)

Pairs of utility functions	Kendall's Tau*	ICC	(95% Confidence interval)
EQ-5D UK tariff versus HUI2	$r = 0.41$	0.51	(0.42–0.59)
EQ-5D UK tariff versus HUI3	$r = 0.37$	0.47	(0.38–0.55)
EQ-5D UK versus Dutch tariff	$r = 0.90$	0.98	(0.97–0.98)
HUI2 versus HUI3	$r = 0.71$	0.74	(0.68–0.78)
HUI2 versus EQ-5D Dutch tariff	$r = 0.40$	0.51	(0.42–0.59)
HUI3 versus EQ-5D Dutch tariff	$r = 0.36$	0.44	(0.35–0.53)

* All statistically significant, $P < 0.01$

**Fig. 1** Scatterplot of utility scores derived with EQ-5D UK tariff and HUI2**Fig. 2** Scatterplot of utility scores derived with EQ-5D UK tariff and HUI3

Impact on the ICER

Mean costs of doing nothing were zero. The mean costs of hearing aid fitting were €1,877. The latter consisted of GP visits (€37), ENT visits (€295) and hearing aid(s) (€1545). All hearing aids were digital, and hearing aids were bilaterally fitted in 83% of the respondents.

This resulted in mean one-year incremental costs of hearing aid fitting versus doing nothing of €1,877. The mean utility gain of 0.01 (sd 0.13), measured with the EQ-5D UK tariff, resulted in a ratio of €286,866 per QALY, with a 95% confidence interval of inferior (higher costs, lower utility) to €47,082/QALY. There was a 36% probability that hearing aid fitting was both more costly and less effective (inferior). The mean utility

Table 4 Change in health state utility after hearing aid fitting and ICER with 95% Confidence Interval (CI) (*n* = 70)

Measure	Mean	Standard deviation	Median	Interquartile range	Minimum	Maximum	ICER ^a €/QALY	(95% CI) €/QALY
EQ-5D UK tariff	0.01	0.13	0.00	0.04	-0.60	0.27	286,866	(inferior ^b -47,082)
EQ-5D Dutch tariff	0.00	0.12	0.00	0.04	-0.60	0.28	647,209	(inferior ^b -61,934)
HUI2	0.07*	0.13	0.08**	0.12	-0.50	0.40	25,337	(19,356–38,012)
HUI3	0.12*	0.18	0.13**	0.22	-0.22	0.60	15,811	(11,664–24,654)

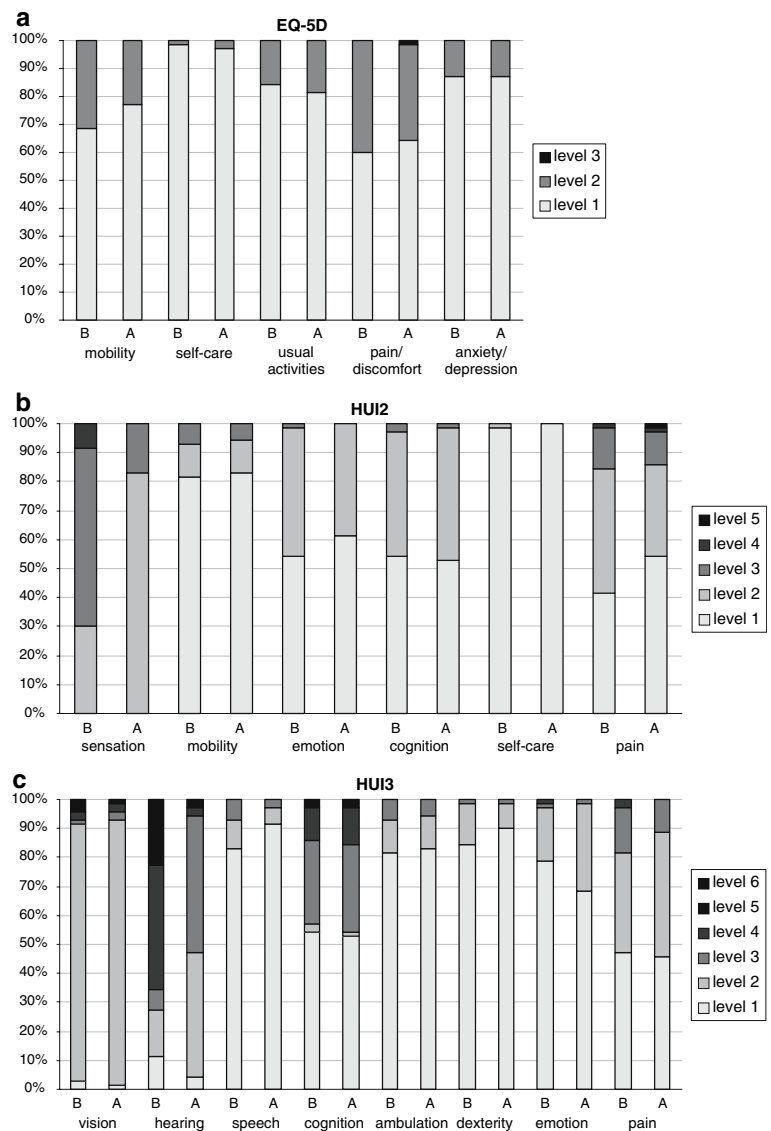
* Statistically significant; Paired *t*-test; *P* < 0.01

** Statistically significant: Wilcoxon Signed-Rank test; *P* < 0.01

^a ICER based on mean scores

^b Inferior means higher costs and lower utility

Fig. 3 Percentage of responses for each level of each attribute before (b) and after (a) hearingaid fitting for EQ-5D, HUI2 and HUI3



gain of 0.003, measured with the EQ-5D Dutch tariff, resulted in an ICER of €647,209 per QALY (95% confidence interval: inferior to €61,934/QALY). There

was a 42% probability that hearing aid fitting was inferior. Applying the HUI2 and HUI3, the ICER was €25,337 per QALY (95% confidence interval: €38,012/

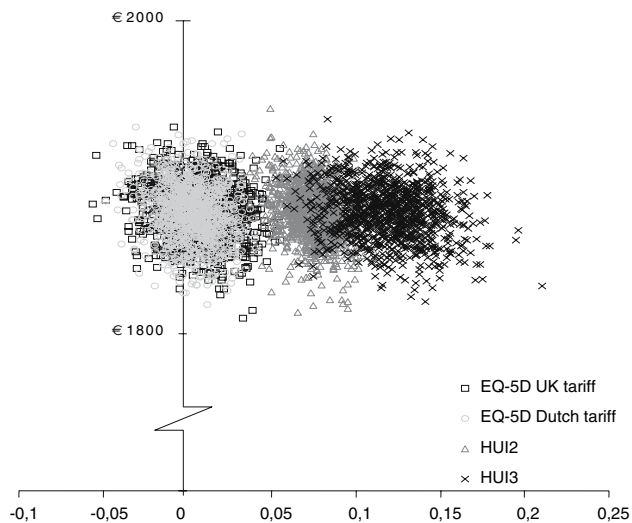


Fig. 4 Incremental cost-effectiveness ratio for hearing aid fitting. Incremental costs (euro) are displayed at the vertical line, incremental effects (QALYs) on the horizontal line

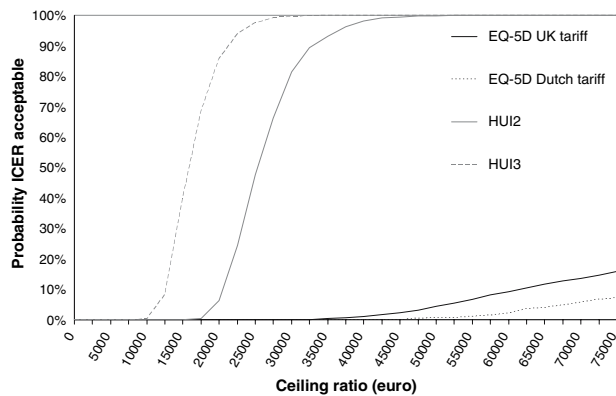


Fig. 5 Cost-effectiveness acceptability curves of EQ-5D UK and Dutch tariff, HUI2 and HUI3

QALY to €19,356/QALY) and €15,811 per QALY (95% confidence interval: €24,654/QALY to €11,664/QALY) respectively. For both measures there was no probability that hearing aid fitting was inferior.

The informal Dutch ceiling ratio of €20,000/QALY [42] implied that hearing aid fitting was only cost-effective when utility was measured with the HUI3. A cost-effectiveness plane with incremental cost and effect pairs for 1,000 bootstrap replications, for all measures, is shown in Fig. 4. Figure 5 shows the cost-effectiveness acceptability curves for all measures.

Conclusion and discussion

Hearing loss affects the exchange of information and therefore affects a person's quality of life [1]. Hearing impaired persons can benefit from using a hearing aid,

since hearing aid use has proven to be effective, it improves social, emotional and communication functions and reduces depression [3, 4].

The results of this study in a relatively healthy population with hearing complaints provide insight in the differences between two widely used utility measures, the EQ-5D and the HUI system.

Regarding practicality, both questionnaires had high completion rates, with the EQ-5D having a higher completion rate than the HUI.

With the EQ-5D, differences were detected in utility by age and sex, indicating construct validity. The HUI3 detected differences by age, but not by sex. Differences between clinically distinctive groups were only detected by HUI3. However, the HUI3 did not confirm our expectation that non-applicants would have higher utility values than first-time applicants. An explanation for this may be that non-applicants had more health problems other than their hearing, as reflected in the ambulation and pain dimensions of the HUI3.

Overall, HUI2 and HUI3 scores were lower than EQ-5D scores and agreement was moderate to poor. Although these measures intend to assess the same construct, namely health state utility, this result was expected as the instruments differ in their underlying assumptions about what constitutes health state utility.

As to responsiveness, only HUI2 and HUI3 measured statistically significant improvement after hearing aid fitting, the EQ-5D UK and Dutch tariff both were not able to capture this effect. Half of the patients (53%) who were fitted with a hearing aid were lost to follow up, either because they had not finished their hearing aid fitting before the end of the study, or because they had their hearing aid fitted at a dispenser not participating in our study. As these patients did not differ from the follow-up group in baseline utility, hearing loss and age, we did not expect this low response rate to influence the results.

The HUI2 and HUI3 change scores resulted in smaller ICERs for hearing aid fitting. Although they were only illustrative, the different ICERs found in the present study clearly show that the choice of a utility instrument in the economic evaluation of hearing aid fitting may heavily influence the cost-effectiveness outcome.

To calculate the ICER of hearing aid fitting we included the total population of respondents considered for hearing aid fitting, regardless of the type of hearing aid fitting. This makes the ICERs found in the present study representative for hearing aid fitting in general in the Netherlands. However, the one year ICERs calculated in the present study were merely illustrative of the impact of different utility scores on the ICER of hearing aid fitting, as the cost-effectiveness of hearing aid fitting has thoroughly been examined by Joore et al [8].

Our results confirm Barton et al [9, 43], who compared the EQ-5D UK tariff and HUI3 before and after hearing aid fitting in a UK sample. However, the change in utility after hearing aid fitting derived with the HUI3 in the present study was twice the change in utility derived with the HUI3 in the study of Barton et al [9]. Respondents in the latter study had a lower baseline utility score and a higher range in utility change after hearing aid fitting, which suggests a somewhat different study population. Our results confirm previous studies that concluded that the EQ-5D lacks sufficient sensitivity to measure the benefit of hearing aid fitting [6, 7, 44]. The lower utility scores and higher responsiveness of the HUI3 in relation to the EQ-5D were also found in a population with visual impairment [22]. Joore et al [8] calculated that hearing aid fitting costs €15,807 per QALY, using the EQ-5D as utility measure. The results of the present study indicate that using the HUI3 as utility measure probably had resulted in a more favourable ICER for hearing aid fitting. Vuorialho et al [45] recently concluded that counseling of hearing aid users to reduce the number of non-users is highly cost-effective, although they were unable to measure any change in utility. As they used the EQ-5D, it is possible that they would have been able to demonstrate favourable costs per QALY when they would have used the HUI3 to measure change in utility.

Three questions arise from the results of the present study: can differences be explained by differences in the measures, are the differences observed between the measures important, and what are the implications of the findings for utility measurement and cost-utility analysis in populations with hearing complaints?

First, differences in utility scores can be explained by differences in the descriptive system and the way the utility scoring function is derived. Regarding the descriptive system, the focus of the EQ-5D on physical, mental and social functioning [46] differs from the ‘within-the-skin’ perspective of HUI, which focuses on the underlying level of impairment. However, this does not explain why the EQ-5D does not measure change after hearing aid fitting, as previous studies have found that hearing aid fitting improves social and emotional functioning [3, 4]. Also, the measures differ in the content and number of attributes, items and levels used, and therefore differ both in the number and in the content of possible health states. It has already been suggested not to use the EQ-5D in relatively healthy populations, given the presence of a ceiling effect [15]. The ceiling effect of the EQ-5D found in the present study is likely to contribute to the differences in responsiveness. When, as found in our study, 44% of the respondents report perfect health at baseline, it is unlikely to find a considerable utility gain from any intervention. Furthermore, since the HUI descriptive system pays

explicit attention to hearing abilities, it is to be expected that in a relatively healthy population with hearing complaints HUI and EQ-5D utility scores differ, and HUI is more responsive. As the HUI3 also pays explicit attention to visual abilities, this may explain why comparisons of EQ-5D and HUI3 in hearing and vision show similar results [22].

There are also differences in the utility scoring functions. Although in general SG (used for HUI) leads to higher scores than TTO (used for EQ-5D) [5], in the present study the EQ-5D scores were considerably higher than the HUI scores. Although different populations do not necessarily yield different results [5], the population sample in which the preferences are measured may also have impact on the differences. In the present study, differences between utilities derived with the UK and Dutch tariff were observed. This difference may be the result of differences in health valuation between people from the Netherlands and the UK, but may also result from the somewhat questionable representativeness of the population sample used to develop the Dutch tariff [27].

Differences also exist in the type of scoring function. EQ-5D uses an additive system, assuming no interaction for preferences among attributes at all. The HUI uses a multiplicative scoring function, with the effect that the loss of utility associated with a particular dimension is dependent on the level of impairment on other dimensions [5]. For example, Barton et al [43] illustrated that hearing impairment (‘unable to hear at all’) has a greater impact on HRQoL as measured with the HUI3 when one has no other health problems (−0.53), than when one also has moderate to severe pain and is unable to see at all (−0.05). It seems rational that persons find their hearing loss a less important aspect of their health state utility when they experience more comorbidity. The multiplicative scoring function of HUI takes this influence of comorbidity into account and seems to be more suitable for modeling utility scores.

Are the differences observed between the measures important? The answer to this question is a clear ‘yes’. The impact of different utility measures on the ICER for hearing aid fitting is of a magnitude that can alter policy decisions and emphasizes that comparisons of QALYs across studies and interventions should be interpreted with caution [12]. The general purpose of a utility measure is to capture the health effects in terms of HRQoL of a policy or program for use in economic analyses. Health economic analyses are a tool to allocate resources in a way that maximizes health (or welfare). In order to use the outcomes of economic analyses for policy decisions, there should however be a clear notion of what should be maximized. The results of the present study show that the potential benefit of an intervention heavily depends on the assumptions of what constitutes health underlying the utility

measure that is used. Beyond the apparent necessity of psychometric quality of a utility measure, an important area for future research is whether societies wish, or should, maximize life expectancy corrected for HRQoL from a more functional perspective as in the EQ-5D, or from a ‘within-the-skin’ perspective as in the HUI. Other studies that have thoroughly examined the psychometric differences between utility measures seem to pass over this important question [11, 13, 19, 21, 24].

What are the implications of the findings for utility measurement and cost-utility analysis in populations with hearing complaints? Generally, it has been recommended that the instrument that is most sensitive to the health states in which one is interested should be selected [5, 12, 47]. From clinical experience it is plain that hearing aid use is effective in alleviating hearing loss and does improve health-related quality of life, but the EQ-5D lacks the sensitivity to capture this improvement. In an otherwise healthy population, HUI3 has proven to be more responsive and therefore more appropriate for evaluating HRQoL in a population with hearing complaints, and is therefore the instrument of first choice in this population.

Acknowledgements Financial support from the Dutch Health Care Insurance Board and the Heinsius Houbolt Foundation is greatly acknowledged. The authors gratefully acknowledge Michelene Chenault for her comments and statistical support. Monique Boymans, Nienke Homans and Silvia Bours are acknowledged for data collection. Preliminary results were presented at the 7th European Federation of Audiological Societies congress, June 19–22 2005 in Göteborg, Sweden, and at the 5th World Congress of the International Health Economics Association, July 10–13 2005 in Barcelona, Spain.

References

- Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *Gerontologist*, *43*(5), 661–668.
- Davis, A. C. (1995). *Hearing in adults*. London: Whurr Publishers Ltd.
- Appollonio, I., Carabellese, C., Frattola, L., & Trabucchi, M. (1996). Effects of sensory aids on the quality of life and mortality of elderly people: A multivariate analysis. *Age & Ageing*, *25*(2), 89–96.
- Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., Rhodes, M. C., Hill, J. A., & DeNino, L. A. (1990). Quality-of-life changes and hearing impairment. A randomized trial. *Annals of Internal Medicine*, *113*(3): 188–194.
- Drummond, M. F., Sculpher, M. J., Torrance, G. W., O’Brien, B. J., & Stoddard, G. L. (2005). *Methods for the economic evaluation of health care programmes*. 3rd ed. New York: Oxford University Press.
- Bess, F. H. (2000). The role of generic health-related quality of life measures in establishing audiological rehabilitation outcomes. *Ear & Hearing*, *21*(4 Suppl): 74S–79S.
- Joore, M. A., Brunenberg, D. E., Chenault, M. N., & Anteunis, L. J. (2003). Societal effects of hearing aid fitting among the moderately hearing impaired. *International Journal of Audiology*, *42*(3), 152–160.
- Joore, M. A., Van Der Stel H, Peters, H. J., Boas, G. M., & Anteunis, L. J. (2003). The cost-effectiveness of hearing-aid fitting in the Netherlands. *Archives of Otolaryngology – Head & Neck Surgery*, *129*(3), 297–304.
- Barton, G. R., Bankart, J., Davis, A. C. & Summerfield, Q. A. (2004). Comparing Utility Scores before and after hearing-aid provision: Results according to the EQ-5D, HUI3 and SF-6D. *Applied Health Economics and Health Policy*, *3*(2), 103–105.
- Bosch, J. L., & Hunink, M. G. (2000). Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Quality of Life Research*, *9*(6), 591–601.
- Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*, *13*(9), 873–884.
- Conner-Spady, B., & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical-care*, *41*(7), 791–801.
- Feeny, D., Wu, L., & Eng, K. (2004). Comparing short form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: Results from total hip arthroplasty patients. *Quality of Life Research*, *13*(10), 1659–1670.
- Kaplan, R. M., Groessl, E. J., Sengupta N, Sieber, W. J., & Ganiats, T. G. (2005). Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. *Medical Care*, *43*(1), 79–87.
- Kopec, J. A., & Willison, K. D. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *Journal of Clinical Epidemiology*, *56*(4), 317–325.
- Longworth, L., & Bryan S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics*, *12*(12), 1061–1067.
- Marra, C. A., Esdaile, J. M., Guh D, Kopec, J. A., Brazier, J. E., Koehler, B. E., Chalmers, A., & Anis, A. H. (2004). A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Medical Care*, *42*(11), 1125–1131.
- Marra, C. A., Woolcott, J. C., Kopec, J. A., Shojania, K., Offer, R., Brazier, J. E., Esdaile, J. M., & Anis, A. H. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Social Science & Medicine*, *60*(7), 1571–1582.
- O’Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. (2003). A view from the bridge: Agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics*, *12*(11), 975–981.
- Oostenbrink, R., Moll, H. A., & Essink-Bot, M. L. (2002). The EQ-5D and the Health Utilities Index for permanent sequelae after meningitis: A head-to-head comparison. *Journal of Clinical Epidemiology*, *55*(8), 791–799.
- Petrou, S., Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics*, *14*(11), 1169–1189.
- Espallargues, M., Czoski-Murray, C. J., Bansback, N. J., Carlton, J., Lewis, G. M., Hughes, L. A., Brand, C. S., & Brazier, J. E. (2005). The impact of age-related macular degeneration on health status utility values. *Investigative Ophthalmology and Visual Science*, *46*(11), 4016–4023.
- Brazier, J., & Deverill M. (1999). A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Economics*, *8*(1), 41–51.
- Pickard, A. S., Johnson, J. A., & Feeny, D. H. (2005). Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research*, *14*(1), 207–219.

25. Rasanen, P., Roine, E., Sintonen, H., Semberg-Kontinen, V., Ryyanen, O. P., & Roine, R. (2006). Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. *International Journal of Technology Assessment in Health Care*, 22(2), 235–241.
26. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, 35(11), 1095–1108.
27. Lamers, L. M., Stalmeier, P. F., McDonnell, J., Krabbe, P. F., van Busschbach, J. J. (2005). Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief. *Nederlands Tijdschrift voor Geneeskunde*, 149(28), 1574–1578.
28. Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. *Health Utilities Index. Pharmacoeconomics*, 7(6), 490–502.
29. The EuroQol Group. (1990). EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.
30. Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang Y, Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. *Health Utilities Index Mark 2. Medical Care*, 34(7), 702–722.
31. Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., & Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care*, 40(2), 113–128.
32. AZOS working group, Final report AZOS. 2006, Health Care Insurance Board: Diemen.
33. Holland, R., Smith, R. D., Harvey, I., Swift, L., & Lenaghan, E. (2004). Assessing quality of life in the elderly: A direct comparison of the EQ-5D and AQoL. *Health Economics*, 13(8), 793–805.
34. GrossPortney, L., & Watkins, M. P. (1993). Foundations of clinical research. Norwalk: Appleton & Lange.
35. Burstrom, K., Johannesson, M., & Diderichsen, F. (2001). Swedish population health-related quality of life results using the EQ-5D. *Quality of Life Research*, 10(7), 621–635.
36. Lubetkin, E. I., Jia, H., Franks, P., & Gold, M. R. (2005). Relationship among sociodemographic factors, clinical conditions, and health-related quality of life: examining the EQ-5D in the US general population. *Quality of Life Research*, 14(10), 2187–2196.
37. Devlin, N., Hansen, P., & Herbison P. (2000). Variations in self-reported health status: Results from a New Zealand survey. *New Zealand Medical Journal*, 113(1123), 517–520.
38. Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. [see comment]. *BMJ*, 316(7133), 736–741.
39. Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.
40. Oostenbrink, J. B., Koopmanschap, M. A., & Rutten, F. F. H. (2004). Manual for costing research (in Dutch). Amstelveen: College voor zorgverzekeringen.
41. Briggs, A. H., Wonderling, D. E., & Mooney, C. Z. (1997). Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics*, 6(4), 327–340.
42. CBO. (1998) Treatment and prevention of coronary heart disease through lowering serum cholesterol levels, National Organisation for Quality Assurance in Hospitals (CBO): Utrecht.
43. Barton, G. R., Bankart, J., & Davis, A. C. (2005). A comparison of the quality of life of hearing-impaired people as estimated by three different utility measures. *International Journal of Audiology*, 44(3), 157–163.
44. Vuorialho, A., Karinen, P., & Sorri, M. (2006). Effect of hearing aids on hearing disability and quality of life in the elderly. *International Journal of Audiology*, 45(7), 400–405.
45. Vuorialho, A., Karinen, P., & Sorri, M. (2006). Counselling of hearing aid users is highly cost-effective. *European Archives of Otorhinolaryngology*, 263(11), 988–995.
46. Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, 37(1), 53–72.
47. Guyatt, G., Jaeschke, R., Feeny, D., & Patrick, D. (1996). Measurements in clinical trials: Choosing the right approach, In B. Spilker (Eds.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (p. 41–48). Philadelphia: Lippincott Williams & Wilkins .