A note on the estimation, the reporting and the use of patient-level QALYs in economic evaluation alongside randomised controlled trials

Andrea Manca, Gerry Richardson, Neil Hawkins, Mark J. Sculpher

Centre for Health Economics, University of York, UK

Abstract

Background: Cost-utility analysis (CUA) is becoming the preferred form of economic evaluation. Its use has been encouraged by guidances issued by several national regulatory bodies such as the National Institute for Clinical Excellence in the UK. Consequently, economic evaluations alongside randomised controlled trials (RCTs) are increasingly being designed to prospectively collect patient-specific resource use and utility data, with the latter used to generate patient-level QALYs. **Problem:** Methodological uncertainty and lack of formal guidance on how patient-level QALYs should be calculated and used in CUA, has resulted in a lack of comparability between evaluations and difficulty in interpretation for decision making. **Methods:** We conducted a systematic literature review of published CUAs to extract information regarding the (a) quality of the reporting of the methods used to generate utilities, (b) methods used to estimate patient-level QALYs, and (c) methods to estimate mean differential QALYs in RCTs. Using data from a recent CUA conducted alongside a large multicentre RCT we evaluated the performances of the methods identified in our literature review to calculate patient-level QALYs and estimate differential QALYs. **Results**: Mean (95% CI) differential QALYs varied considerably from -0.014 (-0.048 to 0.025) to 0.0123 (-0.010 to 0.027) depending on the method adopted in the analysis. **Conclusions:** Analysts should be more explicit in their methodology to enable meaningful comparison between studies. Method of QALYs calculation is crucial and can influence magnitude and direction of results of evaluations. Multivariate regression methods can usefully be applied to address several issues in the estimation of differential QALYs.

Paper presented at the Health Economists' Study Group Winter Meeting
Devonshire Hall, University of Leeds, 8th -10th January 2003

1. Introduction

Economic evaluation conducted alongside clinical trial requires the systematic collection of patient-level resource use and health outcome data over the study period. In cost-utility analysis (CUA) the health outcome of choice is typically the *quality-adjusted life year* (QALY), where patient-level QALYs are calculated on the basis of patients' responses to a generic health-related quality of life questionnaire administered at several time points over the study period. Although CUA has recently been indicated the preferred form of economic evaluation by the UK National Institute for Clinical Excellence (NICE, 2001), no consensus exists on how QALYs should be calculated and analysed in the context of clinical trials.

To explore the extent of the issue, we first conducted a systematic literature review of published CUAs to extract information regarding the (a) quality of the reporting of the methods used to generate utilities, (b) methods used to estimate patient-level QALYs, and (c) methods to estimate mean differential QALYs in RCTs. On the basis of the findings of this systematic review, using data from a recent CUA conducted alongside a large multicentre RCT we then evaluated the performances of the different methods identified in our literature review to calculate patient-level QALYs and estimate differential QALYs.

Data from randomised clinical trials (RCTs) are often analysed on the assumption that the randomisation process ensures *prior* similarity between the trial arms in terms of patients' baseline clinical variables and socio-demographic characteristics. This ensures that any differences in costs and health outcomes observed in the study can be attributed to the intervention being investigated and not to imbalance in the baseline variables between the different treatment groups. However, in CUA alongside RCTs is it often the case that individuals are not stratified on the basis of their baseline utility score, and indeed, it is not infrequent to observe an imbalance in the pre-intervention mean utility score measured at baseline between the different arms of the trial.

This paper argues that regardless whether this being statistically significant or not, the presence of baseline imbalance in the utility score affects the differential QALY results of the study. The mean differential QALY estimate obtained without controlling for this imbalance will be biased, affecting the reliability of the incremental cost per QALY estimate (Pocock *et al*., 2002; Senn, 1989). We will show that, in this circumstance, existing standard techniques to calculate differential QALYs will provide misleading results in terms of differential QALYs and its relative standard error. One solution to this problem is to use a multivariate regression approach, which allows the correct calculation of mean differential QALYs in the trial while controlling for baseline utility value.

The present work is structured as follows. Next section briefly describes the main findings of the systematic review. Some of the main methods used by health economists to estimate QALYs and mean differential QALYs alongside randomised controlled trials are discussed. Section 4 introduces the estimation problems posed by the presence of utility baseline imbalance between trial arms. The results from the application of these methods to a recently completed CUA conducted alongside a large multi-centre clinical trial are reported in section 5.

Finally some simulation results to explore the relevance of the between-patients and within-patients variability in the baseline utility score and their impact on the differential QALY estimation procedure are presented.

2. QALY calculation and reporting in published CUAs: a systematic review

To explore the extent of methodological heterogeneity in QALY calculation and reporting among published CUA studies we conducted an extensive literature review, using the search strategies described in Appendix 1 together with hand searching of specific journals. The search strategies identified 71 papers, of which only 22 were ultimately included in the present review. The remaining papers were rejected being modelling studies, reviews, methods papers, and partial economic evaluation studies.

A data extraction form developed by the authors was used to determine the specific methodology and the level of transparency in the reporting about QALY estimation in each study included in the review. The data extraction form and the results of the review are reported in Table 4. The form is divided in three sections designed to identify information falling under the following headings: (a) study details and preference measure employed; (b) reporting and transparency of QALY estimation; (c) missing utility data and QALY estimation. The main results of the review are described below.

Study details and preference measure employed

The majority of studies collected patient level information on health states or utilities during the study period (n=19), with fourteen papers reporting these data. Three studies used retrospective data collected at the end of the study period to estimate patient level utilities.

The EQ-5D was the most commonly used measure of utility (n=16), with the Health Utilities Index (HUI) the next most commonly employed. The only other study considered used a combination of three measures of utility to estimate scores.

Patients in the study (n=19) were used most often to elicit utilities, while societal values were usually employed (n=19) with a random population sample to determine preferences (n=19).

Reporting and transparency of QALY estimation

Although 18 papers presented baseline utilities or health states, only 5 used these data in their analysis. Moreover several papers did not present utility scores at baseline and follow up (n=9).

Reporting of the variability around utility estimates was poor, with only eight studies reporting a relevant statistic. Similarly for QALY estimation, though the majority of studies (n=18) presented average QALYs in each group only nine studies presented a measure of variability around the QALY estimate.

All the papers considered treated discounting appropriately, but the methodology surrounding utility changes over time and the estimation of QALYs was less

convincing. For utility changes over time, the majority used a linear approach (n=15), though this had to be deduced by the reviewer on most occasions, while two studies used a non-linear approach. The justification for the use of non-linear trend in these studies was not clear and the implications of this assumption were not considered in the sensitivity analysis. Finally, in five studies it was not clear which assumption had been adopted regarding the change in utility values over time.

The methodology used to estimate QALYs varied between papers. The majority (n=12) used the total area under the curve (AUC) approach, but a significant number (n=5) used change from baseline score, while in the remainder of studies (n=7) it was unclear which method had been used. None of the papers considered in this review used a multivariate regression approach to estimate differential QALYs between treatment arms.

Missing utility data and QALY estimation
Finally, few studies dealt with missing data in an adequate manner. Most papers did not acknowledge the missing data problem, though in 12 studies the level of missing data were either reported or could be observed from other data presented. However, only four papers used any form of imputation and only two reported the impact of imputation on their results.

3. Patient-level QALY estimation

In trial-based economic evaluation, patients' utilities are typically elicited inviting patients to answer a specific questionnaire at different time points during the study period. In the simplest scenario, the researcher will have a number of utility measurements over the study period for each individual in the trial to estimate individual-level QALYs. As detailed in the previous section, this is most commonly achieved by applying the area under the curve method (Matthew *et al.* 1990) assuming changes in utility values between consecutive assessments occurs linearly. Figure 1 shows an example of standard QALY calculation for two hypothetical patients receiving two different treatments over a nine-month study period. The utility values for these patients are available at baseline, six months, and nine months (Table 1).

In general terms, a patient-specific QALY can be estimated subdividing the study time into a number of sub-intervals (normally equal to the number of follow-up assessments) and weighting each one of these by a measure representing the patient's utility for that time period. Here we assume for simplicity that the patient's history during the study period is fully observed
The sum each of these quality-adjusted time periods for a patient in the study will generate the patient's QALY, which can be written as

$$QALY = \sum_{j=1}^{n} EQ_j \cdot t_j \qquad (1)$$

where *n* is the number of sub-intervals in which the study period is partitioned, EQ_j represents the utility experienced during the j^{th} interval, and t_j is the duration of the

j^{th} interval expressed in fraction of one year. The shorter the sub-interval the more accurate is the QALY estimation. Unfortunately, for logistic reasons, often questionnaires are administered at discrete time points over the study period with large time intervals between two consecutive assessments. Once individual-specific QALYs have been calculated, the mean differential QALYs in the trial is obtained simply subtracting the mean QALYs for the group of patients in the standard treatment arm from the mean QALYs from the group of patients receiving the new intervention.

4. Differential QALY estimation accounting for baseline EQ-5D imbalance

Notice at this point that, in the example presented in Figure 1, the only difference between the health profiles of the two patients is determined by the value of their pre-intervention baseline utility scores. As a result, despite the patient in the new treatment arm experiences a greater utility improvement compared to the patient receiving the standard treatment, the QALYs calculation supports the adoption of the standard treatment (0.536 vs. 0.494). That is, if we were to ignore the baseline difference in utilities between the two individuals in the study, the differential QALY would be 0.042 in favour of the standard treatment.

This scenario occurs frequently in economic evaluation studies alongside randomised clinical trials, especially in multicentre studies with a pragmatic design where, despite their baseline clinical and socio-demographic variables being not significantly different from the statistical point of view, patients randomised to different interventions may display different baseline utility score. However, the economic evaluation literature seems not to have considered this being an issue in CUA so far. As we have seen in section 2, many CUA studies still fail to report average baseline and follow up utility values for each trial arm, overlooking the presence of imbalance in the baseline utilities.

This section investigates the impact that different estimation methods may have on the quantification of the differential QALYs with its relative standard error, under the assumption that changes in utility values between two consecutive assessments occur linearly.

As detailed in the systematic review, the standard AUC approach is the most commonly used to calculate QALY in CUA, with the differential QALY obtained simply subtracting the average QALY of the control group from that of the intervention group. However, in presence of baseline imbalance in the average utility score between the different arms of the trial, some authors applied the change from baseline approach to estimate QALYs and differential QALYs. The change from baseline method is often used to analyse quality of life data alongside collected RCT at different time points. Here we argue that the change from baseline method is ill suited to the calculation of QALY and differential QALYs in economic evaluation. We argue that average QALYs in each arm of the trial should be calculated using the AUC and suggest the use of the multivariate regression analysis for the quantification of the differential QALY instead.

To illustrate these points let's return now to Figure 1. In graphical terms, the QALY calculation using the mean change from baseline scores means that the QALY of the patient receiving the standard treatment amounts to the area in the

triangle C, and the QALYs for the patient receiving the new intervention would amount to the area given by the sum of the triangles A and C plus the rectangle B. Notice that this method leads to an opposite conclusion with respect to the standard AUC approach, supporting the adoption of the new intervention with the differential QALY equal to the area obtained summing the geometrical shapes A and C.

More importantly, analysing change scores from baseline does not truly control for baseline imbalance in the baseline utility score because of *regression to the mean* problem (Armitage *et al.* 2002; Vickers and Altman, 2001). Where patient-level utilities are collected at several points in time, it is reasonable to expect that an individual's baseline utility value may be negatively correlated, i.e. individuals with low utility values generally improve more than individuals with higher utility scores. In addition, the utility of a patient will be correlated with the mean utility value of the treatment group the individual belongs to.

In this paper we argue that the estimation of differential QALYs and its relative standard error should be carried out within a multivariate regression framework, while controlling for baseline imbalance in the utility values. This is illustrated in Equation 2

$$QALY_i = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot EQ_i^b \qquad (2)$$

where the index i is the patient identifier (i*=1,2,…..N)*, t is a treatment arm dummy variable (0=control; 1= intervention), and EQ_i^b is the patient specific utility at baseline. The coefficients β_0 and β_1 represent respectively the mean QALY in the control arm and the differential QALY, after controlling for any imbalance in the utility at baseline.

In presence of baseline utility imbalance between trail arms, the regression-based approach facilitates an unbiased estimation of the differential QALY (Pocock *et al.*, 2002; Senn, 1989), a more precise estimate of its variance, producing a more efficient than those obtained with other methods (Vickers and Altman, 2001; Vickers, 2001). Notice that if the imbalance exists, this does not need to be statistically significant for the multivariate regression to be usefully employed, as if not controlled for, the baseline imbalance will inflate the underlying error, diminishing the accuracy in the quantification of the sampling uncertainty surrounding the differential QALY. We will come back on this issue in section 6.

In the following section we will compare the three methods described so far, applying them in the context of a recently completed CUA conducted alongside a large multi-centre pragmatic clinical trial.

5. Case study

The EVALUATE trial is the largest comparison of laparoscopic-assisted hysterectomy with standard methods yet undertaken. Details of the results of the clinical and economic analysis from the trial have been published elsewhere (Garry et al. 2002). The time horizon for this initial analysis was one year. Given the design of the trial, in effect two separate comparisons were undertaken. The first estimated the cost-effectiveness of laparoscopic-assisted hysterectomy (ALH)

relative to abdominal hysterectomy (AH) in women for whom the latter would be the conventional procedure of choice. The second estimated the cost-effectiveness of laparoscopic-assisted hysterectomy (VLH) relative to vaginal hysterectomy (VH) in women for whom the latter would be the standard operation.

The economic analysis was undertaken on 1,346 women over a median follow-up of 52 weeks (range 3-52).

The health outcomes for the economic analysis of the alternative forms of hysterectomy were assessed in terms of quality-adjusted life-years (QALYs). QALYs were calculated, for each woman in the trial, on the basis of women's responses to the EQ-5D at baseline, and at up to three points post-operatively (6 weeks, 4 months and one year). Hence, each woman in the trial had a utility at up to four time-points.

Statistical analysis was undertaken using STATA 6.0 (StataCorp, 1999). Due to differential follow-up, 11.68% of patients were subject to administrative censoring. Estimates of mean QALYs, over 12 months follow-up, were, therefore, calculated using regression methods to adjust for censored data (Lin, 2000). Given that the time horizon of the analysis was a year, total QALYs remain undiscounted.

To account for the skewed nature of the data, 95% confidence intervals for the differential QALYs have been calculated using bias corrected non-parametric bootstrapping (based on the 2.5th and 97.5th percentiles) (Efron and Tibshirani, 1993).

Table 2 reports the descriptive statistics for the EQ-5D scores in the EVALUATE study for the two comparisons in the trial, at baseline and at each of the follow up visits. In both comparisons it is evident that there is a little imbalance occurring at baseline between in the EQ-5D score, with a more pronounced disproportion in the ALH vs. AH part of the trial.

Table 3 shows the results of the QALY estimation for this trial using: (a) the standard AUC approach without controlling for baseline EQ-5D scores; (b) the changes from baseline; and (c) the multivariate regression approach described in equation 2 above.

6. Simulation results

In this section we compare the performance of three estimators: (a) the simple (i.e. raw) estimator without controlling for baseline imbalance; (b) the estimator obtained using the change from baseline approach; and (c) the ANCOVA estimator. Their relative efficiency was investigated via simulation. For each ratio tested, 1000 repeats of a 40 subject trial were simulated.

Individual baseline and follow-up observations were simulated as follows.

$$Y_{base} = a + b$$

$$Y_{follow\text{-}up} = a + T + c$$

$a \sim N(0,\sigma^2_{inter})$

$b,c \sim N(0,\sigma^2_{intra})$

where $T \in \{0,1\}$, indicator variable indicating whether subject is in control (T=0) or treatment group (T=1).

The treatment effect was estimated as follows:

1) 'Raw' estimator:

$$\hat{T} = E[Y_{follow\text{-}up}]|T=1 - E[Y_{follow\text{-}up}]|T=0$$

2) 'Change from baseline' estimator:

$$\hat{T} = E[Y_{base}-Y_{follow\text{-}up}]|T=1 - E[Y_{base}-Y_{follow\text{-}up}]|T=0$$

3) 'Regression' estimator:

$\hat{T} = \beta_2$ from the following OLS regression

$$Y_{follow\text{-}up} = \beta_0 + \beta_1 Y_{base} + \beta_2 T + e$$

$e \sim N(0,\sigma^2)$

The variance of each estimator was then estimated as follows:

$$Var(\hat{T}) = (1-\hat{T})$$

The regression estimate of treatment effect had a lower variance than both the 'raw' and 'change from baseline' estimates for all ratios of intra to inter subject variance tested (Figure 2). For ratios of intra- to inter- subject variance above one, the 'change from baseline' estimator actually had a higher variance than the 'raw' estimator. Bearing in mind the likelihood of relatively high intra-subject variability for utility data this is an important consideration. The results indicate that the 'regression' estimator should be used when baseline data are available.

7. Results and discussion

Systematic review

Transparency Issue.

1) Not all studies report baseline and follow up utility data; in a similar way that unit costs should be presented separately from resource use (so that the analysis can be replicated), should baseline and follow up utilities be presented?

2) Several studies did not make their methodology clear

Consistency Issue.

1) Most use EQ5D, several use HUI, one used a combination of 3 utility measures. It is unlikely that these will yield same QALY scores. Similarly, where different techniques used (TTO vs SG) to measure preferences and/or different sources used to determine preferences.

2) Use of different techniques eg to estimate changes in utilities over time, and/or QALYs

Measures of variability

1) Tend to be ignored
2) This has in some cases led to non-statistically significant differences being inappropriately ignored and a cost minimisation analysis (again inappropriately) being conducted.

Missing data

Often ignored (complete case analysis performed), which can lead to substantial biases and erroneous conclusions. Even where missing data acknowledged as a problem it is rarely dealt with adequately.

Comparison of methods

Methods used by health economists to estimate patient level QALYs in clinical trials are heterogeneous. To our knowledge no guidance exists in the economic evaluation literature regarding which approach should be used to estimate QALYs, and particularly in these circumstances. Billingham and colleagues (1999) for instance, present three general approaches to handle changes in health related quality of life between assessments: (a) the *linear change*, (b) *earlier level maintained*, and (c) *change at midpoint*. The linear change method is the most popular approach used by health economists when calculating QALYs and is what we have used in this paper. The 'earlier level maintained" approach can be appropriate in specific circumstances, but we do not know of many applications in health economics literature. The change at midpoint approach in our opinion is equivalent to the linear change, as the area under the curve would be identical in the two cases. Two studies in our review used a non-linear trend used to define the area under the curve. One study fitted *power curves* to model the EQ-5D profiles over the study period. The authors justified this choice arguing that it was appropriate given the time horizon of the study, as the difference in utilities was assumed to exist for the study period but with the magnitude of this difference decreasing over time. However, this approach needs to be tested in the sensitivity analysis. In addition, if adopted extensively the non-linear approach would increase the methodological heterogeneity in QALY calculation, reducing further the transparency of the results by limiting reproducibility of the average results. We believe that once we depart from the assumption of linearity in the change in quality of life between two assessments, it becomes extremely difficult for the researchers to justify the choice of non-linear method used, and although it may be a possibility, its application needs to be investigated further. The degree of departure from linearity may be a difficult element to estimate and it is clearly varying from study to study and from disease are to disease area.

For this reason we compared three methods to calculate QALYs under the assumption that utility changes occur linearly between consecutive periods. We showed that the use of multivariate regression methods is a reasonable approach to handle situations in which imbalance in baseline utility score is present. Although in our case study the differential QALY results do not vary dramatically, it is important to stress that a marginal change in the differential QALY may lead to large changes in the incremental cost-effectiveness ratio, and in some cases to changes in the acceptance decisions regarding the adoption of new health care technologies.

It has been showed that the use of multivariate regression in economic evaluation can facilitate the conduction of standard cost-effectiveness analysis and the exploration of other issues such as patient sub-group analysis (Hoch, Briggs, and Willan, 2002). In this paper we presented another argument for embracing the use of multivariate regression techniques in economic evaluation.

References

Kind, P. The EuroQoL instrument: an index of health-related quality of life. In: Spilker, B, Quality of Life and Pharmacoeconomics in Clinical Trials, Philadelphia: Lippincott-Raven, 1996.

Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. Centre for Health Economics Discussion Paper 172. York: Centre for Health Economics, University of York, 1999

Matthews, JNS, Altman, D, and Campbell, MJ. Analysis of serial measurements in medical research. *British Medical Journal* 1990;300:230-235

StataCorp. 1999 Stata Statistical Software: Release 6.0 College Station, TX: Stata Corporation

Hoch JS, Briggs AH, Willan A. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics* 2002, 11(5): 415-430.

Garry R, Fountain J, Manca A, et al. EVALUATE Hysterectomy Trial. A multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy. *NHS HTA R&D Report, 2002*, forthcoming.

Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000; 1: 35-47.

Efron B, Tibshirani R. *An introduction to the Bootstrap*. New York: Chapman and Hall, 1993

Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* 2001; 323: 1123-4.

Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology* 2001; 1: 16.

National Institute for Clinical Excellence. *Technical Guidance for Manufacturers and Sponsors on making a Submission to a Technology Appraisal* (http://www.nice.org.uk). London: National Institute for Clinical Excellence, 2001

Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; 8: 467-475

Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 2002; 21: 2917-2930

Table 1: EQ-5D scores from a hypothetical RCTs with three assessments (baseline, six months, and nine months)

EQ-5D score	New intervention	Standard treatment
Baseline	0.58	0.68
Six months	0.68	0.68
Nine months	0.75	0.75

Table 2. Health outcomes measured in the EVALUATE trial: responses to the EQ-5D

	Comparison 1				Comparison 2			
	Laparoscopic-assisted hysterectomy (VLH) (n=324)		Vaginal hysterectomy (VH) (n=163)		Laparoscopic-assisted hysterectomy (ALH) (n=573)		Abdominal hysterectomy (AH) (n=286)	
	Mean	Median (IQR)	Mean	Median (IQR)	Mean	Median (IQR)	Mean	Median (IQR)
EQ-5D utilities								
Baseline	0.746	0.760 (0.725 – 1)	0.758	0.796 (0.691 – 1)	0.716	0.760 (0.691 – 0.848)	0.690	0.725 (0.689 – 0.812)
Six weeks	0.875	0.907 (0.812 – 1)	0.852	0.863 (0.76 – 1)	0.832	0.869 (0.76 – 1)	0.833	0.883 (0.76 – 1)
Four months	0.911	0.971 (0.848 – 1)	0.918	0.959 (0.848 – 1)	0.886	0.959 (0.812 – 1)	0.866	0.888 (0.796 – 1)
One year	0.920	1 (0.881 – 1)	0.917	1 (0.861 – 1)	0.897	0.929 (0.848 – 1)	0.892	0.959 (0.822 – 1)

Table 3. Health outcomes measured in the EVALUATE trial: a comparison of methods to estimate Quality-adjusted life years (QALYs) over one year

	Comparison 1		**Comparison 2**	
	Laparoscopic-assisted hysterectomy (VLH) (n=324)	Vaginal hysterectomy (VH) (n=163)	Laparoscopic-assisted hysterectomy (ALH) (n=573)	Abdominal hysterectomy (AH) (n=286)
Method 1 (No adjustment)				
Mean QALYs	**0.895**	0.893	0.868	0.855
Differential QALYs over one year† and (95 % CI)‡	0.0023 (- 0.017 to 0.03)		0.0123 (- 0.010 to 0.027)	
Method 2 (Using Changes from Baseline)				
Mean QALYs	0.150	0.135	0.151	0.165
Differential QALYs over one year† and (95 % CI)‡	0.015 (-0.022 to 0.060)		-0.014 (-0.048 to 0.025)	
Method 3 (Controlling for Baseline EQ-5D)				
Mean QALYs*	0.896	0.893	0.868	0.855
Differential QALYs over one year† and (95 % CI)‡	0.0022 (-0.015 to 0.018)		0.0075 (-0.008 to 0.023)	

* Adjusting for baseline EQ-5D utility

† Laparoscopic-assisted minus standard

‡ 95% non-parametric confidence interval based on 1000 bootstrap replications; lower band = 2.5 percentile; upper band = 97.5 percentile

Table 4 The calculation of QALYs from published CUAs

Category 1. Study details and preference measure employed	**YES**	**NO**	**N/A or Not stated**
1a. Did the authors collect individual-level data on health states or utilities at different time points during the study period?	19 (86%)	3 (14%)	0 (0%)
1b. Were these data reported in the paper?	14 (64%)	8 (36%)	0 (0%)
2. What preference measure was used (eg EQ5D, HUI)?	EQ5D 16 (62%)	HUI 5 (19%)	Other 1 (4%)
3. Who were the health states or utilities elicited from?	Patients in study 19 (86%)	Other 3 (14%)	
4. Whose preferences were used (eg. society, clinician, patients)?	Society 19 (86%)	Other 3 (14%)	
5. From what source were preferences determined (eg random population sample, convenience sample, patients in present study)?	Random pop 19 (86%)	Other 3 (14%)	
6. What technique was used to measure preferences (eg standard gamble, TTO, rating scale)	TTO 14 (64%)	SG 5 (23%)	Other 3 (14%)
Category 2. Reporting and transparency of QALY estimation			
7. Did authors report utility scores at baseline **and** follow up for each arm of the trial?	13 (59%)	9 (41%)	0 (0%)
8. Did the authors present a measure of variability around the utilities?	8 (36%)	14 (64%)	0 (0%)
9. Were average QALYs for each arm of the trial presented (or utility data enabling calculation of average QALYs in each arm)?	18 (82%)	4 (18%)	0 (0%)
10. Was a measure of variability around QALYs presented?	9 (41%)	13 (59%)	0 (0%)
11. What assumptions were used for utility changes over time (eg linear, curve)	Linear 15 (68%)	Curve 2 (9%)	Not clear 5 (23%)
12. Which methodology was used to estimate QALYs (eg change from baseline score, total area under curve)	AUC 12 (55%)	Δ from base 5 (23%)	Not clear 5 (23%)
13. Were effects in future periods discounted if appropriate?	5 (23%)	0 (0%)	17 (77%)

Category 3. Dealing with missing data in QALY estimation			
14. Did the authors report the degree of missing data, or could amount of missing data be deduced from other data presented?	***12*** (55%)	***10*** (46%)	***0*** (0%)
14b. If there were missing data, were data imputed and was imputation method described?	4 (18%)	15 (68%)	3 (14%)
15. Was sensitivity analyses performed to assess the impact of imputation (eg Complete case analysis)?	2 (10%)	20 (91%)	0 (0%)

Appendix 1. Search strategies employed

COST UTILITY ANALYSES & RCTs USING EUROQOL AND HUI AS AN OUTCOME MEASURE

NHS EED Admin Database/Public Database (all years)

s euroqol or eq5d or eq(w)5d or eq-5d
s cost(w)utility or cost-utility
s random$ or rct$ or controlled(2w)(trial$ or stud$) or placebo$ or double(w)blind or single(w)blind or clinical(w)trial$
s s1 and s2 and s3

OHE HEED (April 2002 edition)

ALL DATA: euroqol or eq5d or eq 5d or eq-5d
and
ALL DATA: cost utility or cost-utility
and
ALL DATA: random* or rct* or controlled trial or controlled study or placebo* or double blind or single blind or clinical trial

EconLit/MEDLINE (ARC SilverPlatter) (1995-2002)

#1 euroqol or eq5d or eq 5d or eq-5d
#2 cost utility or cost-utility
#3 random* or rct* or (controlled near2 (trial* or stud*)) or placebo* or double blind or single blind or clinical trial*
#4 #1 and #2 and #3

COST UTILITY ANALYSES & RCTs USING HUI/MUI/QWB/SF-6D AS AN OUTCOME MEASURE

NHS EED Admin Database/Public Database (all years)

s hui or health(w)utilit$(w)index
s (multiattribute or multi(w)attribute)(w)utility(w)index
s quality(2w)wellbeing
s quality(2w)well(w)being
s sf(w)6d or sf6d
s s1 or s2 or s3 or s4 or s5
s cost(w)utility or cost-utility

s random$ or rct$ or controlled(2w)(trial$ or stud$) or placebo$ or double(w)blind or single(w)blind or clinical(w)trial$
s s6 and s7 and s8

OHE HEED (July 2002 edition)

ALL DATA: hui or health utility index or health utilities index
or
ALL DATA: multiattribute utility index or multi attribute utility index or multi-attribute utility index
or
ALL DATA: quality of wellbeing scale or quality of well being scale or quality of well-being scale
or
ALL DATA: sf6d or sf-6d

and

ALL DATA: cost utility or cost-utility

and

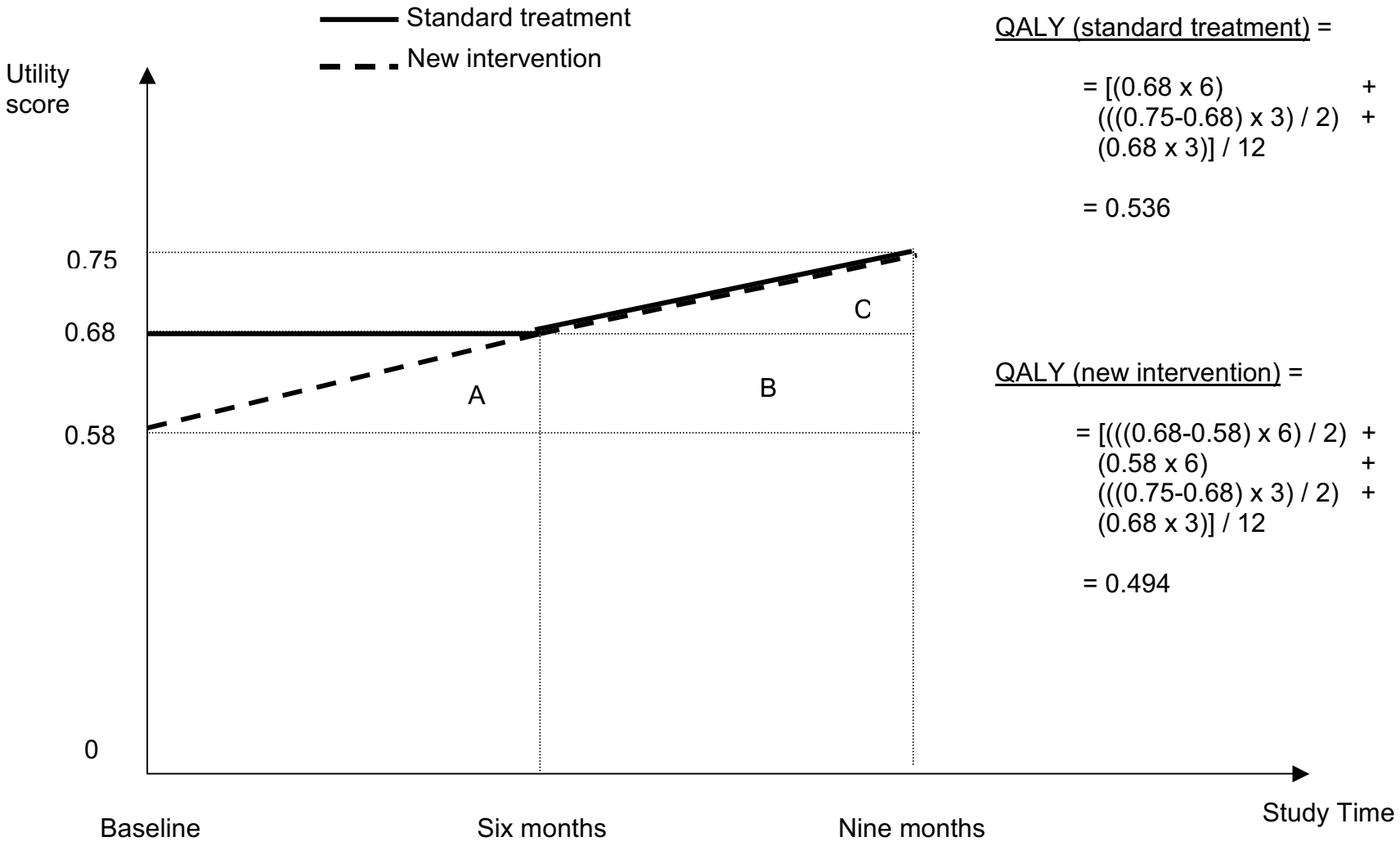
ALL DATA: random* or rct* or controlled trial or controlled study or placebo* or double blind or single blind or clinical trial

EconLit/MEDLINE (ARC SilverPlatter)
(1995-2002)

#1 hui or health utilit* index
#2 (multiattribute or multi attribute) utility index
#3 quality near2 (wellbeing or well being)
#4 sf 6d or sf6d
#5 #1 or #2 or #3 or #4
#2 cost utility or cost-utility
#3 random* or rct* or (controlled near2 (trial* or stud*)) or placebo* or double blind or single blind or clinical trial*
#4 #5 and #6 and #7

END

Figure 1: QALY calculation



QALY (standard treatment) =

= [(0.68 x 6) +
(((0.75-0.68) x 3) / 2) +
(0.68 x 3)] / 12

= 0.536

QALY (new intervention) =

= [(((0.68-0.58) x 6) / 2) +
(0.58 x 6) +
(((0.75-0.68) x 3) / 2) +
(0.68 x 3)] / 12

= 0.494

Figure 2: Comparison of estimators