



QALY Calculation Alongside Randomised Controlled Trials: From the Torch to the Traffic Light

Edwards RT¹, Hounsome B¹, Russell D², Russell I², Williams N³, Linck P¹.

**Paper presented at the 1st Franco-British Meeting in Health Economics
CES/HESG, Paris, 14th – 16th January 2004**

If possible, please print out this paper in colour

¹Centre for the Economics of Health, Institute of Medical and Social Care Research, University of Wales Bangor.

²Institute of Medical and Social Care Research, University of Wales Bangor.

³Department of General Practice, University of Wales College of Medicine.

Abstract

Quality adjusted life years (QALYs) were first developed to shed light on the relative health gain or health gain per £/\$ spent of public health care resources. They provided an index that combined gains in life expectancy with quality of life for use in economic evaluation of health care interventions. Early examples of their use in the UK compared effectiveness in terms of expert clinical prognosis of life expectancy and the health state in which those years would be spent (Williams 1985; Gudex et al., 1990). More recently, the collection of health-related quality of life information (for example, in the form of the EQ-5D measure) during randomised controlled trials has raised the possibility of calculating QALYs based on clinical trial data. Manca and colleagues (Manca et al., 2003) conducted a review of 22 studies reporting QALY figures and found that few health economists adequately report the method by which they calculated QALYs. Manca et al. compared three methods for QALY calculation: (1) area under the curve; (2) change from baseline; and (3) a statistical regression approach. They concluded that a statistical regression approach was superior to the other two methods.

We believe that three serious problems face the health economist responsible for calculating QALYs alongside clinical trials: (1) what to do about missing cost or effectiveness data, e.g. where quality of life is measured at several points over a trial follow-up period; (2) how to allow for any difference at baseline in quality of life between the control and intervention group; and (3) how to generate a valid confidence interval, e.g. using bootstrap replication (Briggs & Gray, 1999).

Our paper challenges the assertion made by Manca and colleagues that the regression method for calculating QALYs is *always* superior to the other possible methods. We discuss solutions to the three problems set out above through a worked example of results from a randomised controlled trial of primary care based osteopathy as compared with usual GP care.

We develop the concept of using a “traffic light” system of colours to aid interpretation of economic evaluation studies (Nixon et al., 2001; Donaldson et al., 2002). We use such a system to display bootstrapped replications in a modified histogram in order to construct a valid confidence interval around a point estimate, showing how it is linked to the cost effectiveness acceptability curve and the cost effectiveness plane.

Finally, in the interest of greater transparency and standardisation in the QALY literature, we offer a brief checklist of steps for the calculation and reporting of cost per QALY figures alongside randomised controlled trials.

1.1 - Introduction

This paper is not about rocket science. It aims to ensure that advances in the use of statistics in health economics, more specifically the economic evaluation of health care interventions, are disseminated effectively amongst health economists and policy makers who look for guidance on the use of limited public health care resources. The concept of the quality adjusted life year (QALY), under a variety of terms, was first developed to shed light on the relative health gain or health gain per £/\$ spent of public health care resources (Klarman et al., 1968; Fanshel & Bush, 1970; Bush et al., 1972; Torrance et al., 1972; Torrance, 1976; Weinstein & Stason, 1977, Drummond et al., 1997). They provided an index that combined gains in life expectancy with quality of life for use in economic evaluation of health care interventions. Early UK examples of their use compared effectiveness in terms of expert clinical prognosis of life expectancy and the health state in which those years would be spent (Williams 1985; Gudex et al., 1990). These authors acknowledged that their evidence base was crude and in need of refinement, but placed faith in the power of the methodology in principle. Almost 20 years on, the National Institute for Clinical Excellence (NICE), responsible for advising the UK National Health Service (NHS) on whether or not to commission and purchase new drugs, therapies and services, still looks to this method to provide insight into the relative value for money in terms of health gain from public health care resources (Forbes et al., 1999; NICE, 2003).

Health care policymakers may mistakenly see QALYs as a panacea for priority setting. To those interested in the mechanics of QALY calculation, they raise more questions than answers. What source of evidence of effectiveness should be used? Whose quality of life values should count? How should quality of life weights be generated? Should all QALYs be of equal weight or should they perhaps be weighted to reflect societal values?

This paper focuses on two broad questions. Firstly, whether philosophically we can compare the results of earlier QALY calculations (which estimate the effect of an intervention on life expectancy and health-related quality of life based on expert clinical prognosis) with present day QALYs based on robust evidence of clinical effectiveness over a short trial duration rather than taking a more crude lifetime perspective.

Secondly, if QALYs are calculated on the basis of evidence from randomised controlled trials (RCT), what is the best method for producing robust generalisable estimates of QALYs gained and cost per QALY gained?

1.2 - Lifetime perspective using expert clinical prognosis, or data from short term clinical trials?

Diagram 1 is the familiar image of QALY gains with and without a treatment. It illustrates the fact that most clinical trials only last a very short period, at most a few years. If an intervention is believed not to affect life expectancy, then one can collect health related quality of life data as part of a trial in order to generate a QALY gain from treatment over no treatment. Even then, there is a decision to be made as to whether or not any improvements in health related quality of life between the intervention group and control group are likely to persist beyond the end of the trial period, although such a decision can be based on the observed data. Where an effect on life expectancy as well as health related quality of life is anticipated, then a judgement has to be made as to the size of that effect, unless all patients have died by the end of the trial period (which would be unfortunate!), and yet again we are left having to make prognoses about the health related quality of life of patients in future years. There is a benefit in collecting health related quality of life data alongside clinical trials for the purpose of calculating QALYs, but accurate patient level data are likely only to be available for a short period relative to the lifetime impact of a health care intervention. Given the role of QALYs as a population level resource allocation guide, they surely need to take a lifetime perspective, if only to avoid implicit biases?

The answer depends on the nature of the intervention being evaluated. If, for example, the intervention is a one-off surgical intervention with a potential effect on life expectancy and/or on quality of life many years in the future, it seems naïve just to use short term clinical evidence of effectiveness from a clinical trial rather than taking a lifetime perspective. However if we are evaluating a treatment for a chronic relapsing problem such as back pain or depression, it seems appropriate to use short-term evidence of clinical effectiveness from a clinical trial, confident that the treatment has a short-term impact on health-related quality of life, but is unlikely to affect life expectancy, and that any benefits to health related quality of life are unlikely to persist far beyond the trial period. Diagram 2 gives an alternative QALY profile for this type of trial. To estimate the QALY gain accurately, several observations of health-related quality of life should be taken during the follow-up period. Ideally this period should extend to the point where the intervention and control curves meet, beyond which the intervention has no further effect. If not, it may be possible to estimate the meeting point by extrapolating both curves. For the purpose of resource allocation we need to be able to compare QALY and cost per QALY figures calculated across different studies. Concerns about our ability to do this have led to criticisms of cost per QALY league tables (Mason et al., 1993). Today, with the emphasis placed by NICE on QALY estimates informing their recommendations, we need to be both mindful of the historical

development of QALYs, and strive towards standardisation of their calculation alongside clinical trials.

2.1 - Three problems for the health economist

Manca and colleagues (Manca et al., 2003) conducted a review of 22 studies reporting QALY figures and found that few health economists adequately report the method by which they calculated QALYs. Manca et al. compared three methods for QALY calculation: (1) area under the curve; (2) change from baseline; and (3) a statistical regression approach. They concluded that a statistical regression approach was superior to the other two methods.

We believe that three serious problems face the health economist responsible for calculating QALYs alongside clinical trials: (1) what to do over missing cost or effectiveness data, e.g. where quality of life is measured at several points over a trial follow-up period; (2) which method to use to calculate QALYs, particularly if there is a difference at baseline in quality of life between the control and intervention group; and (3) how to generate, and disseminate, a valid confidence interval, e.g. using bootstrap replication (Briggs & Gray, 1999).

2.2 - What to do about Missing Cost and Quality of Life Data?

In a randomised controlled trial, if a patient has cost data but no quality of life data, or quality of life data but no cost data, the patient must be omitted from cost-utility calculations. Each patient must be treated as a unit, as costs and effects are likely to be correlated. However, given some information on both costs and quality of life, it is usually better to retain the patient in the analysis by estimating the missing observations from the available data.

This imputation is best done by regression on other observations. For example, if there are four follow-up assessments, one can use all patients with complete quality of life data to obtain a regression equation that predicts the final quality of life value from the baseline and the first three follow-ups. Either use separate equations for intervention and control groups, or include the treatment group as an additional binary predictor in a single equation. If a patient has missed the last follow-up assessment, the baseline and first three follow-up values for that patient can be entered in the equation to predict the missing final value. Demographic information, such as patient age or sex, can be used to refine the prediction if, for example, the treatment has a larger effect on females than on males. However, at risk of stating the obvious, one should not use cost data to predict missing quality of life values. In the same way missing cost data can be imputed

by regressing on the available cost data (and patient characteristics if appropriate), but not on quality of life data or other measures of response to treatment. One exception is that it is seldom useful to impute a missing baseline value, as the remaining follow-up values in isolation contain little useful information about the treatment effect. In the same way, it would be wrong to impute follow-up quality of life values when only the baseline is available.

Two less sophisticated, and usually less desirable, forms of imputation are ‘last observation carried forward’ (so that, for example a missing value at time 3 is estimated by the observed value at time 2), and mean substitution (for example, replacing missing prescription costs by the mean prescription cost in that treatment group).

When missing values have been imputed, it is useful to include a sensitivity analysis that uses only complete data points. A ‘best case, worst case’ approach to missing values can also be used as a sensitivity analysis. For example, in a follow-up of elderly patients whose quality of life is expected to deteriorate, the best case might be ‘last observation carried forward’ and in the worst case all missing patients might be assumed to have died.

2.3 - Which Method of QALY Calculation to Use?

Manca and colleagues (Manca et al., 2003) describe three methods that have commonly been used to calculate QALYs in clinical trials: ‘Area Under Curve’ (AUC), ‘Change from Baseline’ (CfB) and an analysis of covariance approach (‘regression method’).

For each patient the AUC estimate of QALY (quality of life times duration of trial) is the average EQ-5D value over the period of the trial. The estimate of QALY gain in a trial is the difference between the mean of the value in treatment and control groups. For each patient the CfB estimate is the improvement in quality of life compared to its value at the start of the trial, averaged over the duration of the trial follow-up. In contrast to AUC, this allows for chance baseline imbalances when comparing two groups, and usually reduces random variation. The regression estimate uses the AUC values, adjusted for initial differences between treatment groups by regression on baseline values. The resulting estimated difference between the average QALY in the treatment and control groups is intermediate between that obtained by the other two methods. This method will be preferable to the CfB method if there is substantial correlation between

baseline and change scores, such that patients with low initial quality of life scores are likely to improve more than those with high initial scores (Vickers & Altman, 2001).

Manca and colleagues (Manca et al., 2003) recommend using the regression method for calculating QALYs and claim that it is always preferable to the other two methods, illustrating this by simulation from a model. The recommendation is reasonable, as in practice changes in quality of life are usually moderately correlated with baseline values within a treatment group. This is due both to classical ‘regression to the mean’, and to the bounded nature of quality of life measurement scales used in the calculation of QALYs, whereby someone in perfect health cannot improve further.

However, the change from baseline method is a very practical alternative when correlations between improvements in quality of life and baseline quality of life scores are low and sample sizes relatively small. Indeed, if the correlation is small enough, change from baseline is preferable, because a population correlation of zero will not result in a sample correlation of zero. Skewed distributions and non-linear correlations may also affect the performance of the regression method. In non-randomised (observational or quasi-experimental) investigations, where there are systematic differences between groups at baseline, the case for the change from baseline method becomes even stronger.

There are two sources of variation, each with two subdivisions:

- 1). Intra patient variation (‘measurement error’) gives the repeatability of the measure. In this context, it includes (a) short term day-to-day variations in health as well as (b) changes in the response of a patient with unchanged health.
- 2). Inter patient variation (‘real’ differences) includes both differences between (a) the initial state of health of patients as they enter the trial, and (b) variations in how different patients in the same initial state of health respond to treatment.

If inter patient variation is expected to be much larger than intra patient variation, the simplicity and stability of the CfB method should be preferred to the regression method. However, a third type of inter patient variation, correlated with the initial state of health of the patient, has a similar

effect to intra patient variation. The model used by Manca and colleagues (Manca et al., 2003) did not include the second type of inter-patient variation.

The change from baseline method can easily be bootstrapped in order to generate a confidence interval. We have adopted this approach with respect to the case study presented. We are aware, however, that our observed intra group correlations are sufficiently large that a regression method of QALY calculation should be also be considered. Choice of the regression method for calculating QALYs, along with a wish to generate a bootstrapped confidence interval, necessitates a recalculation of the regression equation for every iteration of the bootstrap (where at least 1000 replications is usual practice). This, we believe, would be preferable to alternative model-based estimates, which usually assume normality and linear correlations, and sometimes also assume that cost and QALY gain are uncorrelated.

2.4 - How to generate confidence intervals around QALYs or cost per QALY estimates?

The need for confidence intervals around clinical outcomes in a trial has been well accepted (Gardner & Altman, 1989). It is fair to say that health economists are only now appreciating and acknowledging the importance of statistical robustness in what was, until recently, a very new methodology of economic evaluation. Health economists with statistical expertise have argued over the best way to construct a confidence interval around an incremental cost effectiveness ratio. Cost-effectiveness ratios are often highly skewed especially if, as in the following example, cost and effect differences between treatments are not large. Many model-based approaches make the unrealistic assumption that costs and effects are independent. It has, therefore, been argued that a non-parametric bootstrap replication is the best way to construct such a confidence interval (Briggs & Gray, 1999).

3.1 - Case study of osteopathy as addition to usual primary care for back pain

Spinal pain is common and costly to health services and society. Management guidelines have encouraged primary care referral for spinal manipulation, but the evidence base is weak. More economic evaluations alongside pragmatic trials have been recommended. Our aim was to assess the cost-utility of a practice-based osteopathy clinic for sub-acute spinal pain.

The design of our study was a cost-utility analysis alongside a pragmatic single-centre randomised controlled trial. Its setting was a primary care osteopathy clinic accepting referrals from fourteen neighbouring practices in North West Wales. Patients with back pain of 2 to 12 weeks duration were randomly allocated to treatment with osteopathy plus usual GP care (n = 86) or usual GP care alone (n = 101). Costs were measured from a National Health Service (NHS) perspective. All primary and secondary health care service utilisation by patients for back pain and all other reasons was collected from GP notes for the six month study period. We calculated Quality Adjusted Life Year (QALY) gains based on EQ-5D responses from patients in the trial, and then calculated cost per QALY ratios. Confidence intervals were estimated by non-parametric bootstrapping using a new “traffic light” method for displaying bootstrapped replications. Osteopathy and usual GP care was more effective but resulted in more health care costs than usual GP care alone, although neither difference reached statistical significance. The point estimate of the incremental cost per QALY ratio was £3,560 (80% CI £540, £77,070). When spine related costs alone were considered, the point estimate of the incremental cost per QALY ratio was £3,040 (80% CI £1000, £35,100). Rigorous multi-centre studies are needed to assess the generalisability of this approach, but it is also necessary to take account of sampling variation.

The traffic light system for characterising a confidence interval around a cost per QALY estimate is a system for assigning colours to different quadrants of the cost effectiveness plane, and showing corresponding colours on a bootstrap histogram. This shows how the confidence interval is constructed. We build on attempts by health economists and clinicians to use colour to clarify the dissemination of the findings of economic evaluations, and thus to aid health care decision making.

Nixon and colleagues have argued that the method for summarising the results of economic evaluations is not well established (Nixon et al., 2001). They reviewed three methods for summarising the results of economic evaluations: narrative summaries, permutation matrices and the cost effectiveness plane. They suggested that the cost effectiveness plane is potentially very useful as it shows the magnitude of the differences in costs and effectiveness, as well as their direction. However, they argued that this approach is somewhat technical in nature, lacks common reference points that can be used across different economic evaluations, and has not been widely used in systematic reviews. They therefore proposed a new method for summarising the results of systematic reviews of economic evaluations, using shading in a hierarchical

decision matrix to provide an instant visual summary of the findings and the likely actions for the decision maker. They conceded that the idea of using colour in clinical decision making is not new, citing its use in guidelines on cholesterol treatment (Kwaliteitsinstituut, 1997). Soon after, in a paper by Donaldson and colleagues, on the limitations of cost effectiveness analysis to address allocative and technical efficiency issues, the BMJ editors added a traffic light system of colours to aid dissemination and interpretation of their figure (Donaldson et al., 2002; Donaldson, 2002).

3.2 - Using the “traffic light” approach to derive a bootstrap confidence interval around a cost per QALY result for our worked example

In our case study cost-utility analysis of osteopathy in addition to GP usual practice for back pain, QALYs were calculated for each patient from the EQ-5D scores at baseline, two and six months. Missing EQ-5D values were imputed using regression models. There was no difference in mortality between the two groups. The method used to calculate QALYs was the change from baseline method (Manca et al., 2003), and changes in utility values between consecutive assessments were assumed to occur linearly. Because by six months there was no difference between the mean EQ-5D scores of the two groups, it was reasonable to assume that osteopathy had no effect on quality of life beyond the duration of the trial.

A confidence interval for the cost per QALY ratio was estimated using the non-parametric bootstrap method (Efron & Tibshirani 1993). A 1000 replication bootstrap generated 1000 estimates of the cost per QALY ratio, which were plotted on the cost-utility plane (Figure 1). A stacked histogram (Figure 3) of the cost per QALY estimates was plotted for the values in the "trade off" quadrants, coloured yellow and blue for NE and SW quadrant points respectively. Values in the other two quadrants did not have meaningful ratios, so those in the SE quadrant (lower cost and QALY gain) were grouped into a single green bar at the left hand edge of the histogram; those in the NW quadrant (higher cost and QALY loss) into a single red bar at the right hand edge. A confidence interval was determined by excluding the appropriate proportion of cases at either end of this histogram, with an iterative correction for points in the blue “trade off” quadrant. Finally, a cost-utility acceptability curve (Figure 2) was plotted comparing the ceiling value of the cost per QALY ratio with the probability that the intervention gives good ‘value for money’. This analysis was repeated for costs attributed to spinal pain as a sensitivity analysis, but is not presented here.

The median bootstrap cost-utility estimate was £3,760 per QALY gained, which was similar to the cost utility point estimate described earlier. The 80% confidence interval was £540 to £77,070. Thus a decision maker prepared to pay at least £77,070 to gain a QALY would be 90% confident that osteopathy should be introduced; one not prepared to pay as much as £540 would be 90% confident that it should not be introduced. If the cost acceptability threshold is between the confidence interval limits, neither choice can be made with the required degree of certainty. The decision maker should wait for further evidence. However, the usual convention is to require 95%, rather than 90%, confidence before introducing a new treatment. An 85% confidence interval included cases from the North West quadrant, where the intervention was never worthwhile. If the confidence level was increased to 90% the interval also included cases in the South East quadrant where there were cost savings and QALY gains. Such an interval effectively represents complete uncertainty: regardless of what value is placed on a QALY, the intervention may be either desirable or undesirable. Despite the apparently favourable cost utility point estimate, this trial alone does not provide enough evidence to conclude that osteopathy is a cost effective addition to usual care. This is not a surprising conclusion, as osteopathy was neither significantly more effective nor significantly less costly than usual care. Thus conclusions from cost-utility analyses need to include estimates of statistical uncertainty as well as sensitivity analyses, as isolated point estimates of cost per QALY ratios can be misleading.

A cost-utility acceptability curve was plotted (Figure 2). This curve did not originate at zero probability when the ceiling ratio was £0, because of the proportion of bootstrap estimates that fall in the South East quadrant of the cost utility plane, or the green bar at the left edge of the cost per QALY histogram. Similarly, the asymptote of the slope never reached 1.0, because of the proportion of bootstrap estimates that fall in the North West quadrant, or the red bar in the histogram.

Apart from the anomalous South West quadrant values, the bootstrapped cost-utility acceptability curve is effectively a cumulative frequency curve of the distribution represented by the histogram. Thus they are two ways of representing the same information, and the confidence intervals derived above can also be obtained from the curve. The advantage of the histogram is that it is a clearer and more familiar representation, and distinguishes between points in the North East and South West quadrants, which have different implications for health service budgets.

We believe this “traffic light” approach to constructing a confidence interval around an estimate of QALY gain or cost per QALY gain or, for that matter, any incremental cost effectiveness ratio greatly aids communication with both fellow health economists and, more importantly, non-health economists as to how the confidence interval was obtained and its credibility.

4.1 - Proposed checklist for QALY calculation alongside clinical trials

A number of useful general checklists exist for the conduct, publication, and critical appraisal of economic evaluation studies (e.g. Drummond, 1997). We were, however, unable to find one specific to the calculation of QALYs alongside clinical trials.

Checklist at start of trial

1. Gather clinical consensus on whether intervention is likely to have an effect on life expectancy, and whether any observed improvements in health related quality of life are likely to be sustained and for how long with respect to life expectancy.
2. Make the period for the collection of quality of life data as long as necessary to detect plausible differences.
3. Collect patient-specific quality of life data as frequently as practicable throughout the trial.
4. Collect as much patient specific cost data as necessary to detect plausible differences on all service use.

Checklist for analysis of QALY gains from clinical trial data

1. Check whether you have complete cost and effectiveness data for each patient in the control and intervention groups and remove patients for whom there is either no cost data or no effect data.
2. If there is some missing effect data from, for example, one of three data collection points in the trial, impute missing data. Impute missing cost data from other cost data.
3. Check equivalence of reported health-related quality of life at baseline. If different, relate change in quality of life to baseline quality of life scores independently for intervention and control groups.
4. Check equivalence of cost data at baseline if cost data has been collected, for example, for the six months prior to the start of the intervention period (through retrospective review of GP records). If different, relate mean patient costs for intervention period to mean patient costs for pre-trial period for intervention and control groups.

5. Select method of calculation. If there is no significant difference between groups in effect at baseline, choose change from baseline method. If there is a large correlation between change in quality of life score and baseline scores for each group, think about a regression approach.
6. Present estimates of QALY gain and cost per QALY gain based on trial data.
7. Conduct bootstrap replications to generate a confidence interval.
8. Choose width of confidence interval depending on distribution of plotted bootstrap cost-utility ratios across the cost-utility plane.
9. If effectiveness of intervention is thought likely to extend beyond the duration of the observed clinical trial period, run sensitivity analysis of minimum and maximum duration of effect on life expectancy and health-related quality of life.
10. If constructing cost per QALY estimates, state year of costs and source of costs and how they were obtained.
11. Apply discount rate where appropriate to costs and possibly to QALYs.
12. If there are serious outliers of cost or effect, re-run the analysis to check as to whether their exclusion changes the conclusions of the analysis, particularly their effect on the bootstrapped confidence interval.
13. Check whether the cost per QALY figure depends on context in the sense that resources required to switch from control to intervention depend on local infrastructure (Donaldson et al., 2002).
14. Consider and report robustness of cost per QALY estimate in light of statistical significance of clinical outcomes of trial.

References

- Briggs AH, Gray IS. (1999) Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 3(2).
- Bush J, Fanshel S, Chen M. (1972) Analysis of a tuberculin testing program using a health status index. *Socio-Economic Planning Sciences* 6: 49-68.
- Craig N, Sutton M. (2002) Addressing the contra-indications to cost effectiveness analysis in health care *www.bmj.com*.
- Donaldson C, Currie G, Mitton C. (2002) Cost effectiveness analysis in health care: contra-indications. *BMJ* 325:891-4.
- Donaldson C. (2002) Contra-indications to cost effectiveness analysis: responses to Craig and Sutton and von Elm *www.bmj.com*.
- Donaldson C. (2002) Re: Use of colour in the Effectiveness-cost matrix *www.bmj.com*
- Drummond MF, O'Brien B, Stoddart GL, Torrance GW. (1997) *Methods for the economic evaluation of health care programmes*. Second Edition. Oxford: Oxford University Press.
- Fanshel S, Bush J. (1970) A health status index and its application to health services outcomes. *Operations Research* 18(6): 1021-66.
- Forbes RB, Lees A, Waugh N, Swingler RJ. (1999) Population based cost- utility study of interferon beta-1b in secondary progressive multiple sclerosis. *BMJ* 319:1529-1533.
- Gardner MJ, Altman DG. (1989) *Statistics with confidence*. London: BMJ.
- Gudex C, Williams A, Jourdan M, Mason R, Maynard J, O'Flynn R, Rendall M. (1990) Prioritising waiting lists. *Health Trends* 3: 103-8.
- Klarman H, Francis J, Rosenthal G. (1968) Cost-effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care* 6(1): 48-54.
- Kwaliteitsinstituut voor de gezondheidszorg (1997) Guideline cholesterol. *CBO*.

Manca A, Richardson G, Hawkins N, Schulpher MJ. A note on the estimation, the reporting and the use of patient-level QALYs in economic evaluation alongside randomised controlled trials. Paper presented at the Health Economists' Study Group Meeting, January 2003.

Mason J, Drummond M, Torrance G. (1993) Some guidelines on the use of cost-effectiveness league tables. *BMJ* 306: 570-572.

National Institute for Clinical Excellence (NICE), (2003) Website: <http://www.nice.org.uk>.

Nixon J, Khalid K, Kleinjen J. (2001) Summarising economic evaluations in systematic reviews: a new approach. *BMJ* 322: 1596-8

Nixon J. (2002) Use of colour in the Effectiveness-cost matrix *www.bmj.com*

Torrance GW, Thomas W, Sackett D. (1972) A utility maximization model for evaluation of health care programs. *Health Services Research* 7(2): 118-33.

Torrance GW. (1976) Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences* 10(3): 129-36.

Vickers AJ, Altman DG. (2001) Analysing controlled trials with baseline and follow up measurements. *BMJ* 323:1123-1124.

Weinstein M, Stason W. (1977) Foundations of cost-effectiveness analysis for health and medical practices. *New Eng J Med* 296(13): 716-21.

Williams A. (1985) Economics of coronary artery bypass grafting *BMJ* 291:326-329.

Acknowledgements

The authors would like to thank Mark Sculpher and colleagues for permission to quote the HESG paper prepared by Manca et al., 2003.

The authors would also like to thank Andrew Briggs for use of his cost-effectiveness Excel macro.

Diagram 1: Typical QALY profile for an RCT where health gains are long-term

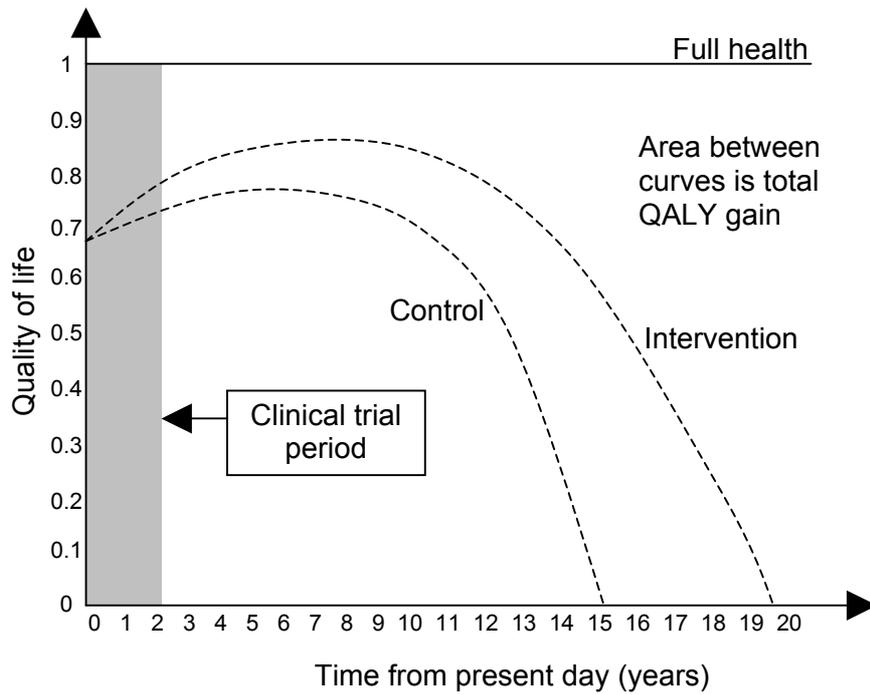


Diagram 2: Typical QALY profile for an RCT where health gains are short-term

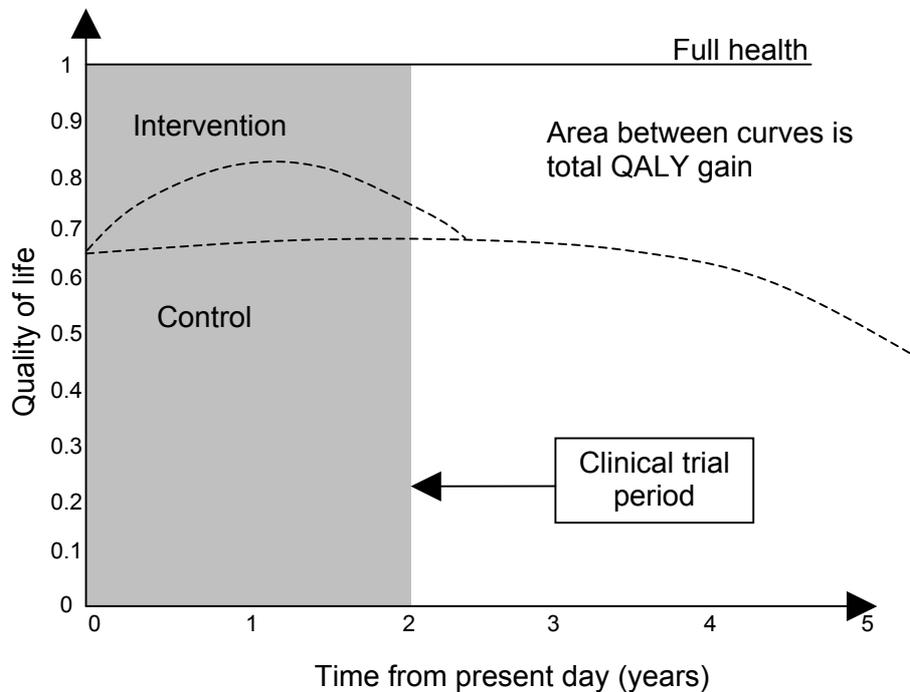


Figure 1: The cost-utility plane

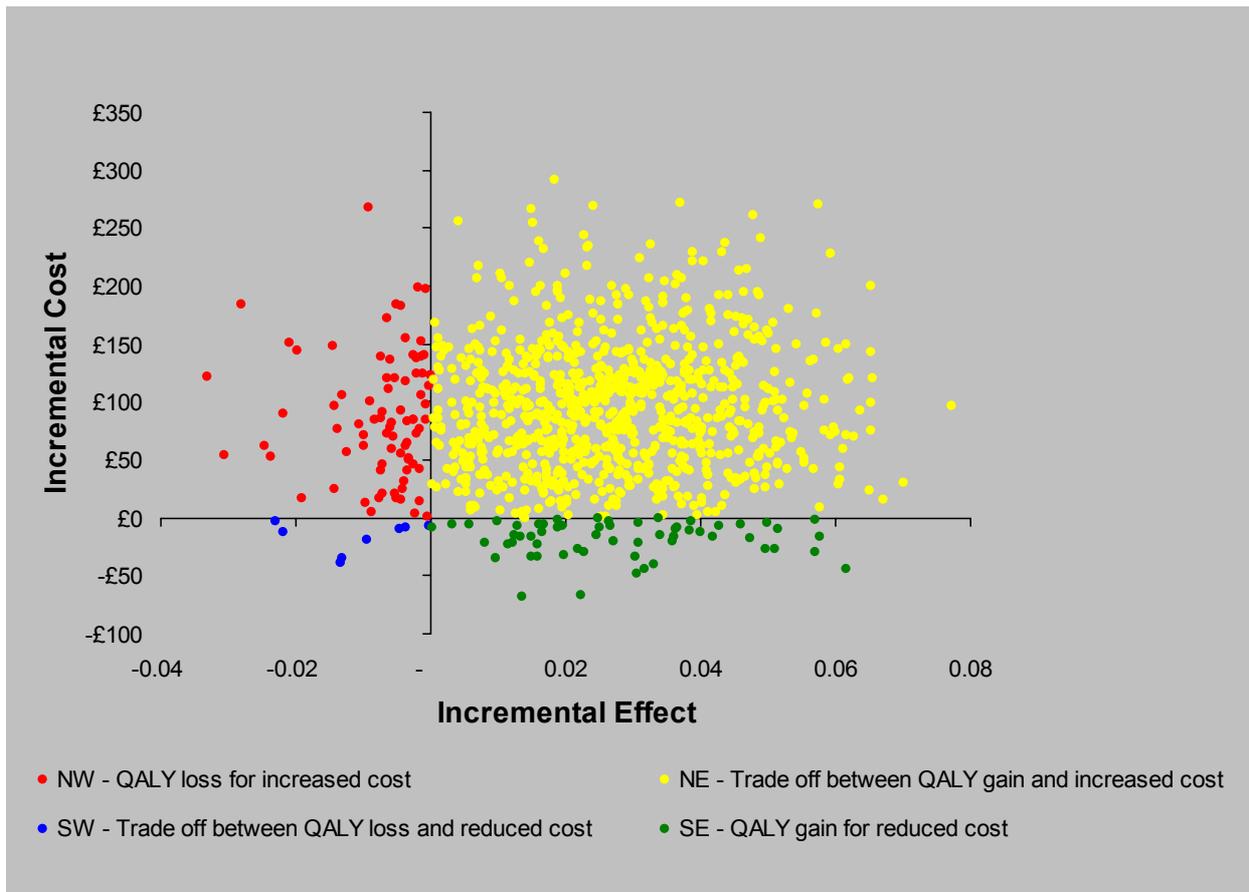


Figure 2: Cost-utility acceptability curve

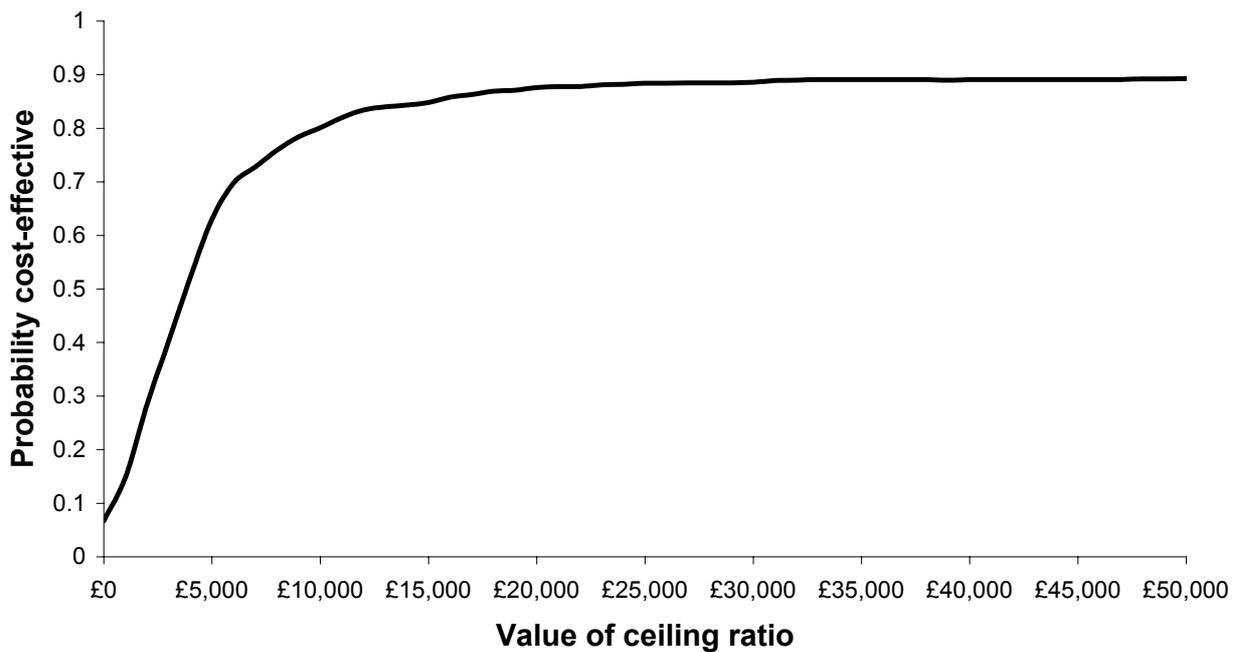


Figure 3: Traffic light histogram showing frequency of bootstrapped cost per QALY values

