WORK IN PROGRESS. PLEASE DO NOT QUOTE OR REFER TO WITHOUT THE PERMISSION OF THE AUTHORS

A Comparison of QALYs and Healthy-Years Equivalents and their Relationship with Individuals' Preferences

Mark Sculpher
Marco Barbieri

Centre for Health Economics, University of York

Paper presented at Health Economics Study Group meeting, Oxford, January 2001.

---

1. Introduction

The quality-adjusted life-year (QALY) is now widely used in economic evaluation in health care.[1] During the period of growth in the use of the QALY, there has been a methodological tension between the view that the QALY is simply an index of health, and the perspective that it reflects individuals' preferences. Although the early papers on the QALY concept focused on its index characteristics,[2], the QALYs tenuous link with preferences is frequently cited as a disadvanatge.[3,4]

The QALY, as conventionally used in economic evaluation, is derived on the basis of four stages. Firstly, health profiles (prognoses), following from alternative forms of patient management, are decomposed into a series of discrete health states. Secondly, these health states are valued independently. Sometimes a distinction is made during the valuation process between temporary and chronic health states. Thirdly, the period of time in a given health state is multiplied by its value. QALYs are calculated by summing these products over the duration of the patient's survival (or over the time horizon of the study). This process may involve discounting QALYs generated in future years. Finally, uncertainty can be incorporated into this process by calculating a series of QALY profiles and attaching a probability to each. Expected QALYs are calculated by multiplying each QALY profile by its relevant probability and summing across all the profiles. In trial-based economic evaluations, this expectation is calculated across patients.

For this 'standard QALY' to reflect individuals' preferences, it is necessary for treatments generating more QALYs to be preferred by individuals' over those producing fewer QALYs. However, this link is based on some important assumptions about the individual's utility function.[4-7] Firstly, risk neutrality with respect to life-years needs to hold, which implies a utility function for life-years which is linear in life-years. This is an assumption which is at odds with the evidence.[8,9] The use of the standard gamble to generate the quality weights for QALYs does not remove the need for this assumption because, although the quality weights may reflect individuals' risk attitude, the life-years component will not, as it remains in natural units. Furthermore, doubts exist about whether quality weights based on the standard

gamble adequately reflect individuals' risk preferences given the predictive limitations of expected utility theory.[10,11]

The second assumption, is that life-years and quality are mutually utility independent, implying that a quality weight for a given health state will be the same irrespective of the time in the health state. The third assumption is that of constant proportional trade-off, which means that an individual is willing to sacrifice a constant proportion of their remaining period of survival to acquire a given improvement in health status, whatever the absolute number of life-years that remain. The evidence on this assumption is mixed. [5,7,12] The final assumption is required when QALYs are used to reflect preferences over a sequence of health states over time; namely, that individuals' valuations of a given health state are independent of the health states that precede or follow it. This additive independence is central to the use of valuation systems such the Health Utilities Index and EuroQol (EQ)-5D, but virtually all cost-utility analyses require this assumption. Again, the evidence on this assumption is mixed, with some studies refuting it,[13,14] but others showing results consistent with it.[15]

The healthy-years equivalent (HYE) was developed as an alternative measure of benefit for cost-utility analysis which sought to form a stronger link with individuals' preferences by avoiding some of the assumptions of the standard QALY.[3,16] The HYE retains the QALY's purpose of incorporating the impact of a technology on both quantity and quality of life, and can be defined as 'the hypothetical combination of years in a state of full health, which is equal, in terms of the individual preferences (utility), to the patient's current projected lifetime health profile'.[3] Although, as originally described, the HYE was based on an elicitation process using a two-stage standard gamble exercise, [16] the above definition indicates a close relationship between the HYE and the time trade-off (TTO) – indeed, it has been argued that the two-stage standard gamble and the TTO are theoretically equivalent.[17,18] The TTO can be used to elicit a HYE by describing a profile of health states over time to a respondent and asking for the period of time in good health equivalent to that profile. The TTO is conventionally used to measure preferences under conditions of uncertainty. To elicit HYEs with the TTO, however, it would be possible to present the health state profiles with some information about the probabilities attached to those profiles. Johannesson has likened this to the eliciting a certainty equivalent HYE (an *ex ante* HYE) that is equal, in terms of preferences, to a risky health profile.[19] In many ways, this is similar to undertaking a willingness to pay exercise where respondents are asked to value an intervention with a risky health profile.

The use of the TTO to elicit *ex ante* HYEs in this way has the potential advantage of generating a measure of benefit for cost-utility analysis which has a closer relationship to individuals' preferences than the standard QALY. However, the practicality of the valuation process needs to be considered. Although simplification in the health profile descriptions will be necessary, it is not clear whether the presentation of profile information and data about risks will over-burden respondents.

This paper reports on a study to explore the practicality and value of the TTO-based *ex ante* HYE in the clinical context of the surgical treatment of menorrhagia. The specific objectives of the analysis were, firstly, to assess whether individuals could complete the TTO exercise based on descriptions of a risky health profile; secondly, to compare the results of the cost-utility analysis based on HYEs with that using standard QALYs; and, thirdly, to consider how the standard QALY and the TTO-based HYE compare with other information respondents provided regarding their treatment-related preferences.

2. Methods

2.1 Clinical context

Menorrhagia, or excessive regular menstrual bleeding, affects about 22% of otherwise healthy women in the UK[20] and 5% of women aged 30-49 years consult their GP with heavy bleeding each year.[21] First line therapy is usually with drugs, but these often fail to ameliorate symptoms adequately, and many women are referred to hospital for the consideration of surgical treatment. The mainstay of surgical treatment for menorrhagia has traditionally been hysterectomy, undertaken using open surgery on the abdomen.

Since the early 1990s, there has been a widespread diffusion of minimal access surgical treatments which remove the lining of the uterus whilst leaving the organ in place. About 10,000 of these procedures were undertaken in the UK in 1993, 82% of which were transcervical resections of the endometrium (TCRE) using diathermy.[22] Randomised trials comparing abdominal hysterectomy (AH) and TCRE show that there is a clear trade-off between the two treatments, with TCRE resulting in shorter hospitalisation and a less severe convalescence, but failing to improve symptoms adequately in around 20% of women.[23-25] Furthermore, studies have shown that a proportion of women have strong preferences about the attributes of the two forms of treatment.[26,27]

2.2 Standard QALYs

A cost-utility analysis of AH and TCRE has been published based on a two-year time horizon.[28] In brief, the study used a decision analytic framework within which to synthesise data from a randomised trial,[24,29] with health state preference data elicited outside the trial. A sample of 60 women recently referred to hospital with menorrhagia were presented with descriptions of 5 health states incorporated into the model: unsuccessfully treated menorrhagia; convalescence following TCRE; convalescence following AH; premenopausal following recovery from a successful TCRE; and premenopausal following recovery from AH. Health state values were elicited using the TTO, with all states except the two relating to convalescence (which were treated as temporary states) valued as chronic states over each woman's life expectancy. To estimate QALYs, these health state values were incorporated into the decision model, placed in composite profiles and translated into treatment-specific expected QALYs using probabilities taken from the trial. Hence the derivation of QALYs in this analysis was undertaken in the conventional way: health states were valued by individuals who were asked to imagine being in that state for a fixed period; the values are then used to quality-weight health states over quite different time periods and in particular sequences; and a composite QALY was calculated by aggregating these quality-weighted time periods.

2.3 TTO-based ex ante HYEs

A separate exercise was mounted to elicit HYEs, against which the standard QALY results would be compared. Given the importance of not overburdening women with too many valuation tasks, it was not considered appropriate to use the same sample of women to obtain TTO-based HYEs as was used to elicit health state values for the standard QALY estimates. Therefore, a further sample of women was identified in a second centre. In order to identify a sample of women with very similar characteristics, the same process was used to recruit women into the study.[28] A sample of 63 women, identified from GP referral letters received by the hospital, agreed to participate in the valuation study.

The interview involved using the TTO to elicit values for health profiles in terms of *ex ante* HYEs. Two *ex ante* profiles were presented to women, one for AH and one for TCRE. These included a clear time dimension running from initial surgery until the menopause, and then until death. These profiles included estimates of the risks associated with therapy: the risk of operative death (for both treatments) based on estimates for AH which were assumed to apply to both treatments,[30] and of re-treatment (for TCRE) based on the results of the clinical trial.[24,29] The two scenarios were based partly on the synthesis of the health state

descriptions used in the health state valuation exercise upon which the standard QALYs were based, and partly on additional information from the same sources used to develop the original descriptions. Appendix I reproduces the two profiles used to elicit *ex ante* HYEs.

As the profiles had a time dimension lasting until death, they had to be 'customised' for each woman's life expectancy. As for the health state valuation exercise, life expectancy was assumed to be 60 years (for women aged between 20 and 29 years); 50 years (for those aged between 30 and 39 years); 40 years for those aged between 40 and 49 years; and 30 years (for those aged between 50 and 59 years). The profiles assumed the menopause would occur with approximately 35 years of life remaining.

On being presented with the profiles, the women were asked to rank them. The *ex ante* HYEs for the two profiles were then elicited using the TTO. As for the health state valuation study, the 'converging ping-pong' method was used to avoid anchoring bias.[31] The process of eliciting *ex ante* HYEs using the TTO is the same as that used to elicit health state values to estimate QALYs, the only difference being that the period of time in imperfect health considered comparable to a lifetime profile is itself the HYE estimate, and is not then transformed to a value on a 0 to 1 scale.

2.4 Comparing QALYs and HYEs

One objective of the analysis was to compare the results of the cost-utility analysis when standard QALYs and TTO-based *ex ante* HYEs are used as the measure of benefit. To do this, the results of the analysis using standard QALYs required adjustment in order that they relate to the same time horizon as was used to elicit HYEs using the TTO, namely a woman's lifetime. The two-year analysis is extrapolated until the menopause assuming that women will continue until the menopause in the health state in which they are located at two years; the only differences in resource consumption between the two treatment groups are that, of patients retaining their uterus, 74% undergo cervical cytology every five years[32] and 10% use hormone replacement therapy,[29] compared to 0% and 17%, respectively, of women who have undergone a hysterectomy; and the average age of the menopause is 51 years.[33] Extrapolation until death assumes that, after the menopause, women move into a good health state for the remainder of their life valued at 1; consumption of health service resources is the same for all patients over that period; and the mean age of death is 80 years.

When HYEs are elicited from women, their time preference rates are reflected within the values they provide. In comparing QALYs and HYEs, therefore, expected QALYs associated with AH and TCRE are presented with a range of discount rates.

2.5 Comparing QALYs and HYEs with women's treatment preferences

The rationale for HYEs is that they more adequately reflect individuals' preferences than standard QALYs. To assess whether this was the case in the context of the surgical management of menorrhagia, patient-specific QALYs and HYEs were compared to other information obtained from women on their treatment-related preferences. Patient-specific QALYs were calculated for each of the 60 women in the health state valuation exercise by separately incorporating each woman's health state values into the decision model.

In addition to characterising women's preferences in terms of health state values for *ex ante* HYEs, a range of stated preferences about treatments and treatment attributes was collected from women using a questionnaire. Full details of the questionnaire are provided elsewhere.[27] Three parts of the questionnaire were used in the comparison with QALYs and HYEs. The first part listed a series of 10 characteristics of surgical treatment for menorrhagia (see Appendix II), and women were asked to rate the importance of these characteristics on a 4-point scale from 'very important' to 'not important'. The second part

described two options for treatment, one representing AH and one TCRE. Women were asked to indicate which, if any, they would prefer, and to rate each of them on a visual analogue scale. In the third part, women were asked to indicate whether they had strong preferences for or against any particular treatments and, if so, to name them. Women's responses to these parts of the questionnaire have been used to assess the consistency of the patient-specific QALYs and *ex ante* HYE estimates with these more descriptive treatment preferences. The following analyses have been undertaken.

(a) Characteristic groups. Based on women's responses to the questions about the importance of the various characteristics of treatment for menorrhagia, two 'characteristic groups' were defined. If women considered the characteristics typical of TCRE as important, they were placed in the TCRE characteristic group; if they felt the characteristics typical of AH were important, they were put into the AH characteristic group. A score of 4 was assigned to the treatment characteristics that women considered "very important", a score of 3 to the treatment characteristic considered "of some importance", a score of 2 to the "of little importance" ones, and a score of 1 to the treatment characteristics considered "not important". Mean scores for the treatment characteristics related to TCRE and AH were separately calculated. Women were placed into the TCRE group if their mean score for attributes typical of TCRE (e.g. short as possible length of hospital stay; treatment that does not remove the womb) was higher than the mean score for attributes typical of AH (e.g. not having to worry about contraception in the future; wanting to stop periods for good). Women were placed in the AH characteristic group if their mean score for AH was higher than their mean score for TCRE.

(b) Choices about treatment options. Women's responses to the choice between the two treatment options described in the questionnaire are compared to their patient-specific QALYs and *ex ante* HYEs. Consistency would require that QALYs and HYEs for a given treatment would be higher for women who indicated that she would select the option describing that treatment in the questionnaire.

(c) Specific treatment preferences. Women's responses to questions about positive and negative preferences for actual treatments are also compared to their QALYs and *ex ante* HYEs. Due to small numbers, both positive and negative preferences are grouped as being for hysterectomy or for other treatments. Although much depends on women's prior information about the characteristics of treatments, it would be reasonable to expect QALYs and *ex ante* HYEs to be higher (lower) when a woman stated a strong positive (negative) preference for that treatment.

3. Results

3.1 Samples of women providing values

Table 1 describes the women interviewed in the two valuation exercises. The table shows that the two groups were very similar apart from the length of time during which they had experienced menorrhagia. For this characteristic, women in the health state valuation sample (for standard QALYs) had experienced their symptoms for a median period of 24 months compared to 12 months in the *ex ante* HYE sample.

3.2 Health state values and ex ante HYEs

Table 2 presents details of the health state values elicited from the 60 women in the valuation sample. Mean (95% CI) and median (range) values are presented for each of the five health states which were described to women. The results show that the ordering of the chronic health states is the same in terms of mean and median values. As expected, the chronic health state scenario valued lowest by women was menorrhagia which, in terms of mean values, women were prepared to trade 50% of their future life expectancy to avoid.

The chronic health state scenario describing health after an AH, but prior to the menopause, was valued most highly.

Table 3 summarises the *ex ante* HYEs elicited from the second sample of women using the TTO on the basis of the single risky health profiles presented for each treatment (see Appendix I). All 63 women were able to complete the valuation exercise. The table shows that the mean and median HYEs were higher for AH than for TCRE, although this difference was not statistically significant as shown by the overlapping 95% confidence intervals. The mean HYE for AH, for example, indicates that, on average, women considered 34.84 years in good health to be equivalent to the risky profile following AH assumed to take place over an average period of 44.76 years.

3.3 Comparison of standard QALYs and ex ante HYEs

In order to compare the *ex ante* HYEs with standard QALYs, it is necessary to use the decision model to estimate patient-specific QALYs using each woman's individual health state values. The results of this calculation are also shown in Table 3. With standard QALYs, time preference is incorporated by the analyst at the point of synthesis within the model whereas, with HYEs, the respondent's time preference is reflected in the values they provide. To facilitate some comparison, therefore, the standard QALYs are presented using three different discount rates to express time preference: 6%, 2% and 0%.

The forms of benefit measure show the same direction of benefit: with both QALYs and HYEs, AH is shown to generate greater benefit for the average women. A comparison of the two forms of benefit measure in terms of absolute benefit depends on the discount rate used to calculate QALYs. The UK Department of Health's recommended rate of 2% per annum[34] produces absolute QALY values approximately 85% that of the *ex ante* HYEs. These results, and those for QALYs calculated with no discounting, suggest that, in this context, the discount rate that would equate QALYs and HYEs would lie somewhere between 0% and 2%. However, the differential in benefit between the two treatments is greater for the *ex ante* HYEs than with QALYs.

The decision model used for the cost-utility analysis based on standard QALYs provided an estimate of expected costs, over two years, of £794 for TCRE and £1,139 for AH.[28] Extrapolated over a woman's life expectancy, this translates, using a 6% discount rate, into expected costs of £816 and £1,162, respectively – a differential cost of £346. Hence an incremental cost per unit of benefit on the basis of standard QALYs (with a 2% discount rate) would be £372. This compares with an incremental ratio, using *ex ante* HYEs, of £156, the lower value reflecting the higher differential in benefit between the two treatments with HYEs. On the basis of standard QALYs calculated using women's mean QALYs, AH would probably be considered cost-effective when the incremental cost-effectiveness ratios of other interventions are used as a form of comparison. Although what constitutes an 'acceptable' incremental ratio is even more difficult to assess with HYEs than with QALYs because of the lack of applied studies using HYEs, it is likely that the modest cost per additional HYE would result in AH being considered cost-effective on the basis of both forms of benefit measure.

3.4 Comparing QALYs and HYEs with women's treatment preferences

Three separate analyses were undertaken to assess the consistency between patient-specific standard QALYs (based on incorporating each woman's health state values into the decision model) and *ex ante* HYEs.

3.4.1 Characteristic groups

As described above, women were allocated to the AH or TCRE characteristic group on the basis of how important they rated attributes of treatment. Based on their responses to the

questionnaire, therefore, the characteristic group to which a woman is assigned is expected to reflect their preferences regarding treatment attributes. Based on an analysis of individual responses, Table 4 shows how the assigned characteristic group compares with individuals' treatment preferences as implied by patient-specific QALYs and HYEs. For women in the health state valuation sample, 20/59 fell into the AH characteristic group. For these women, 20 (90%) had higher patient-specific QALYs for AH than for TCRE, on the basis of their personal health state values. Of the 30/59 women who were assigned to the TCRE characteristic group, however, only 8 (31%) had higher patient-specific QALYs for TCRE than for AH. This results in an overall inconsistency rate for standard QALYs of 29/59 (49%).

For the sample of 63 women from who *ex ante* HYEs were elicited,19/62 fell into the AH characteristic group. For these women, 12 (63%) had higher HYEs for AH than for TCRE. Of the 43/62 women who were assigned to the TCRE characteristic group, however, 12 (28%) had higher HYEs for TCRE than for AH, with 17 (40%) having equal HYEs for the two treatments. This results in an overall inconsistency rate for HYEs of 17/62 (27%).

3.4.2 Choices about treatment options

In completing the questionnaire, women were presented with descriptions of AH and TCRE, and were asked to indicate a preference. Women's responses to this question are compared to their patient-specific QALYs and *ex ante* HYEs in Table 5. For the sample of women who provided health state values, 29/59 (49%) said that they preferred AH as described in the questionnaire and, of these, 26 (87%) had higher patient-specific QALYs for AH than for TCRE. Of the 21/59 (36%) women who chose TCRE as described in the questionnaire, 90 (38%) had higher patient-specific QALYs for TCRE than for AH. This gives an overall inconsistency rate of 15/59 (25.4%).

Of the sample of women from whom *ex ante* HYEs were elicited, 24/58 (41%) chose AH as described in the questionnaire and, of these, 17 (71%) had higher HYEs for AH than for TCRE. Of the 24/58 (41%) who chose TCRE as described in the questionnaire, 7 (29%) had higher HYEs for TCRE than for AH, giving an overall inconsistency rate of 11/58 (19%).

3.4.3 Specific treatment preferences

When completing the questionnaire, women were given the opportunity to specify any treatments that they had strong preferences either to receive (positive preference) or not to receive (negative preference). Given the small numbers involved, positive and negative preferences have been split into those for hysterectomy and those for 'other treatments'. A group-level comparison of these stated preferences with patient-specific QALYs and TCRE is shown in Table 6. For women who stated a strong positive preference for hysterectomy, mean standard QALYs and mean HYEs were higher for AH than for TCRE, although the differential in HYEs between AH and TCRE is greater than the differential in standard QALYs between treatments. In both the sample of women providing health state values for standard QALYs and that from whom HYEs were elicited, those women having a strong positive preference for a treatment other than hysterectomy had higher mean patient-specific benefits (QALYs and HYEs respectively) for AH than for TCRE.

For women who had strong negative preferences for hysterectomy, both mean standard QALYs and mean HYEs were higher for TCRE than for AH. In the sample of women providing health state values who had a strong negative preference for an other treatment, mean patient-specific QALYs were also higher for TCRE than AH. In the sample of women providing HYE values with a strong negative preference for another treatment, however, mean HYEs were higher for AH than for TCRE.

4. Discussion

4.1 The practicality of TTO-based HYEs

The strength of the standard QALY is that the value judgements necessary to synthesise multi-dimensional measures of outcome into a single benefit measure are (or should be) made explicit. When presented with the results of a QALY-based economic evaluation, a decision maker can accept or reject its conclusions; but in rejecting them, alternative value judgements will have to be discussed and presented. The QALY, therefore, can serve a valuable role within resource allocation: as one tool in the decision maker's armamentarium for purposes of resource allocation; as a means of making judgements about the synthesis of multi-dimensional outcomes explicit; as a broad-brush means of comparing outcomes, as well as costs, across programmes and disease areas; and as a way of initiating a consideration of the economic characteristics of health care technologies. As such, the QALY is a complement, rather than an alternative, to the disaggregated approach to economic evaluation that has been advocated.[35] However, this 'decision making perspective' on the value of QALYs is quite different to the view that QALYs are a means of incorporating individuals' preferences into resource allocation.

For some, this pragmatic rationale for QALYs will be sufficient to justify their use in applied studies. Others will argue that the necessary assumptions to link the QALY with individual preferences are too heroic for them to be useful, and there is a need for a more preference-based measure of benefit for CUA. The HYE is but one of several measures that have been advocated for this purpose, including the risk-adjusted QALY#26740 and, with particular reference to individuals' social preferences, the person-trade-off.[36] However, the practicality of their use in applied studies has not been extensively studied. The first objective of this paper was to consider the practicality of using one alternative measure of benefit for CUA: the *ex ante* HYE elicited using the time trade-off. In the context of the economic evaluation of alternative treatments of menorrhagia, the process of eliciting patients' values using this method was found to be feasible: all women provided values and appeared to understand the task. It should be emphasised that the risky profile presented to women did not cover all possible eventualities following treatment. In order to avoid overloading women with too much information to assimilate, it focused on the main risks (death and, for TCRE, treatment failure) and the likely sequence of health states over time. However, simplification is an inevitable part of benefit valuation for CUA; for example, for the standard QALY, not every conceivable health state is defined and valued.

Although this study suggests that the valuation burden of the TTO-based HYE is manageable, the analytical burden should also be considered. Standard QALYs have the advantage of being a flexible benefit measure. For example, the decision analytic model is a popular framework for QALY-based analysis, where parameters such as probabilities and durations in health states can be varied, and the implications for cost and benefit results assessed, without altering the health state values. Because scenarios for HYEs include more information, the HYE estimates have analytical parameters locked within them, and an assessment of the robustness of an analysis to changes in these parameters is impossible unless revised HYEs are elicited. This may suggest that, if TTO-based HYEs are to be used widely in economic evaluation, it would require a rather different approach to the timing of benefit valuation work within the overall study. If a RCT is the major source of outcome information for an economic assessment, the valuation data necessary to facilitate a QALY-based study would usually be collected within the trial using an instrument such as the EQ-5D. If HYEs are to be estimated, the valuation exercise would have to await the results of the trial, so that descriptive profiles could be developed to include all relevant information. For CUA based on decision modelling, the estimation of QALYs would require the valuation data to be incorporated into the model, with QALYs being a major outcome of the exercise. With

HYEs, however, the model would have to be used to generate the information to go into the descriptive scenarios, which would then be used to elicit HYEs.

4.2 Comparing study results with QALYs and HYEs

The second objective of the paper was to consider, in the clinical area of the management of menorrhagia, whether the use in cost-utility analysis of TTO-based HYEs, instead of standard QALYs, made a different to the conclusions of the study. The results of the earlier modelling work reached a clear conclusion that AH was more costly, over a two-year period, than TCRE.[28] Furthermore, based on conservative extrapolation assumptions, this finding is likely to remain over a lifetime time horizon. The standard QALY analysis showed that, based on women's average health state values, AH generated more QALYs over 2 years; and, again, conservative assumptions would suggest that this would also be true over a lifetime. The *ex ante* HYE data reported here also indicate that, over a lifetime time horizon and on the basis of women's average values, AH would generate greater benefit than TCRE. Therefore, in the *direction* of differential benefit between AH and TCRE, the two forms of benefit measure show consistent results.

Which treatment is considered more cost-effective, however, depends on whether the additional cost of AH is justified by the incremental benefit. In the case of standard QALYs, an incremental cost per QALY gained with AH of £372 (assuming a 2% annual discount rate on benefits) compares favourably with other funded therapies in the NHS. Hence, the intervention would probably be considered cost-effective by decision makers. In the case of HYEs, the differential benefit is greater between AH and TCRE (2.22 versus 0.98, with a lifetime time horizon and a 2% annual discount rate for QALYs), so the incremental cost per unit of benefit gained in less (£156). In interpreting the incremental ratio with HYEs, the lack of similar data from other studies using the benefit measure is a constraint. However, it seems unlikely that an additional cost of £156 per additional HYE would be considered too great for AH to be funded. If this is true, the standard QALYs and HYEs would probably lead to the same conclusion: based on average patient values, AH is a more cost-effective surgical intervention for menorrhagia than TCRE.

The limited number of applied studies using HYEs precludes any assessment of the extent to which this result is generalisable to other clinical contexts. Only one other study has compared the cost-effectiveness of alternative interventions using *ex ante* HYEs, based on the TTO, and standard QALYs. Cook *et al*[37] considered the context of three alternative treatments for gallstone disease: open and laparoscopic cholecystectomy and extracorporeal shockwave lithotripsy (ESWL). Using information from interviews with patients and patient questionnaires, a series of health state scenarios was developed. These scenarios took a 'partial *ex ante*' perspective in that the treatment-related scenarios included information on the risk of operative mortality. On the basis of interviews with 96 members of the general public, TTO exercises were undertaken to value each scenario. The authors then compared the differential QALYs associated with the alternative treatments and their aftermaths with differential HYEs elicited form a partial *ex ante* perspective. The authors found differences between the two measures in terms of the absolute value of benefit generated by the alternative treatments. There were also some differences in the overall cost-effectiveness results: in terms of standard QALYs, laparoscopic cholecystectomy was found to be a dominant treatment strategy but, using *ex ante* HYEs, ESWL was found to be more effective and more costly. However, the fact that, from the *ex ante* HYE perspective, ESWL had an incremental ratio of Aus$2,228,000 suggests that it would not be considered good value for money using standard QALYs or *ex ante* HYEs. Hence, as for the study reported here, the conclusions were insensitive to the type of benefit measure used.

4.3 Comparing benefit measures with women's treatment preferences

The rationale for the HYE is that it avoids some of the assumptions that are necessary to link the standard QALY to individuals' preferences. The fact that no studies have sought to test this hypothesis empirically, partly reflects the perennial problem in the benefit valuation field that there is no gold standard measure of health-related preferences. The analysis presented here has attempted to relate standard QALYs and TTO-based HYEs with other expressions of treatment-related preferences that women have provided. The treatment-related preference data that women were asked to provide in the questionnaire was only ordinal in nature. That is, in indicating which attributes of treatment were important to them, which, if any, of the therapies that were described to them they would choose and which treatments which they were aware of they would like to undergo, they provided no information about the strength of their preferences. However, if HYEs are indeed more able to reflect individuals' preferences than standard QALYs, an assessment of the consistency between these benefit measures and the preference information women supplied in the questionnaire would be expected to support this.

The data presented here provide some indication that the HYEs elicited from women were more consistent with the treatment-related preferences supplied in the questionnaire. In terms of both the treatment preferences inferred by the importance women attached to treatment attributes, and the treatments women chose on the basis of the descriptions offered to them, there was less inconsistency with the preferences implied by patient-specific HYEs than with patient-specific QALYs. However, some important caveats are necessary. Firstly, the small numbers of women in the study limits statistical power. Secondly, one attribute that was not included in the questionnaire related to the possible need for repeat surgery following TCRE, and this may have influenced the responses and the numbers of patients allocated to the two characteristic groups. Thirdly, due to the need to avoid over-burdening respondents, health state values (for standard QALYs) and HYEs were elicited from two separate samples of women. This precluded any within-subject assessment of the link between preferences and the alternative benefit measures. Given these limitations, the results regarding the consistency of QALYs and HYEs with preferences should be treated cautiously, and further research in this area is required.

In conclusion, the use of the TTO to elicit *ex ante* HYEs on the basis of descriptions of risky profiles was found to be feasible. In the context of a cost-utility analysis of two surgical treatments for menorrhagia, however, the use of HYEs produced similar results to those with standard QALYs. Tentative support was provided for the hypothesis that *ex ante* HYEs are more consistent with individuals' preferences than standard QALYs.

References

1. Chapman RH, Stone PW, Sanberg EA, Bell C, Neumann PJ. A comprehensive league table of cost-utility ratios and a sub-table of 'Panel-worthy' studies. *Medical Decision Making* 2000;20:451-467.

2. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England Journal of Medicine* 1977;296:716-721.

3. Mehrez A, Gafni A. Quality-adjusted life years, utility theory and healthy-years equivalents. *Medical Decision Making* 1989;9:142-149.

4. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Social Science and Medicine* 1989;28:299-308.

5. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Research* 1980;28:206-224.

6. Johannesson M. Quality-adjusted life-years versus healthy-years equivalents - a comment. *Journal of Health Economics* 1995;14:9-16.

7. Bleichrodt HJM. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. *Medical Decision Making* 1997;17(1):21-32.

8. McNeil BJ, Weichselbaum R, Pauker SG. Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine* 1978;299:1397-1401.

9. Eraker SA, Sox HC. Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making* 1981;1:29-39.

10. Schoemaker PJH. The expected utility model: its variants, purposes, evidence and limitations. *Journal of Economic Literature* 1982;20:529-563.

11. Richardson J. Cost utility analysis: what should be measured? *Social Science and Medicine* 1994;39:7-21.

12. McNeil BJ, Weichselbaum R, Pauker SG. Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine* 1981;305:982-987.

13. Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. *International Journal of Technology Assessment in Health Care* 1996;12:151-162.

14. Kuppermann M, Shiboski S, Feeny D, Elkin EP, Washington AE. Can preference scores for discrete states be used to derive preference scores for an entire path of events? *Medical Decision Making* 1997;17(1):42-55.

15. Mackeigan LD, O'Brien BJ, Oh PI. Holistic versus composite preferences for lifetime treatment sequences for type 2 diabetes. *Medical Decision Making* 1999;19:113-121.

16. Mehrez A, Gafni A. The health-years equivalent: how to measure them using standard gamble approach. *Medical Decision Making* 1991;11:140-146.

17. Buckingham K. A note on HYE (healthy years equivalent). *Health Economics* 1993;11:301-309.

18. Weinstein M, Pilskin J. Perspectives on Healthy-years Equivalents: HYEs:What are the issues. *Medical Decision Making* 205-206.

19. Johannesson M. The ranking properties of healthy-years equivalents and quality-adjusted life-years under certainty and uncertainty. *International Journal of Technology Assessment in Health Care* 1995;11:40-48.

20. Gath D, Osborn M, Bungay G, Iles S, Day A, Bond A *et al*. Psychiatric disorder and gynaecological symptoms in middle aged women: a community survey. *British Medical Journal* 1987;294:213-218.

21. Effective Health Care. *The Management of Menorrhagia. Effective Health Care Bulletin*

No. 9. Leeds: Nuffield Institute for Health, University of Leeds, 1995.

22. Overton C, Hargreaves J, Maresh M. A national survey of the complications of endometrial destruction for menstrual disorders: the MISTLETOE study. *British Journal of Obstetrics and Gynaecology.* 1997.;104:1351-1359.

23. Gannon MJ, Holt EM, Fairbank J, Fitzgerald M, Milne A, Crystal AM *et al*. A randomised trial comparing endometrial resection and abdominal hysterectomy for the treatment of menorrhagia. *British Medical Journal* 1991;303:1362-1364.

24. Dwyer N, Hutton J, Stirrat G. Randomised controlled trial comparing endometrial resection with abdominal hysterectomy for the surgical treatment of menorrhagia. *British Journal Obstetrics Gynaecology* 1993;100:237-43.

25. O'Connor H, Broadbent M, Magos AL, McPherson K. Medical Research Council randomised trial of endometrial resection versus hysterectomy in management of menorrhagia. *Lancet* 1997;349:897-901.

26. Coulter A, Peto V, Doll H. Patients' preferences and general practitioners' decisions in the treatment of menstrual disorders. *Family Practice* 1994;11:67-74.

27. Sculpher MJ, Dwyer N, Browning J, Horsley S. A survey of women's preferences regarding alternative surgical treatments for menorrhagia. *Health Expectations* 1998;1:96-105.

28. Sculpher MJ. A cost-utility analysis of abdominal hysterectomy versus transcervical endometrial resection for the surgical treatment of menorrhagia. *International Journal of Technology Assessment in Health Care* 1998;14:302-319.

29. Sculpher MJ, Dwyer N, Byford S, Stirrat GM. Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health-related quality of life and costs two years after surgery. *British Journal of Obstetrics and Gynaecology* 1996;103:142-149.

30. Dicker RC, Greenspan JR, Strauss LT, et al. Complications of abdominal and vaginal hysterectomy among women of reproductive age in the United States. *American Journal of Obstetrics and Gynaecology* 1982;144:841-848.

31. Mohide EA, Torrance GW, Streiner DL, Pringle DM, Gilbert R. Measuring the wellbeing of family caregivers using the time trade-off technique. *Journal of Clinical Epidemiology* 1988;41:475-482.

32. Brown J, Sculpher MJ. Economics of screening programmes to prevent cervical cancer. *Contemporary Reviews in Obstetrics and Gynaecology* 1993;5:221-229.

33. Luoto R, Hemminki E, Topo P, Uutela A, Kangas I. Hysterectomy among Finnish women: prevalence and women's own opinions. *Scandanavian Journal of Social Medicine* 1992;20:209-212.

34. Department of Health. *Policy Appraisal and Health*. London: Department of Health, 1995.

35. Freemantle N, Mason J. Not playing with a full DEC: why development and evaluation committee methods for appraising new drugs may be inadequate. *British Medical Journal* 1999;318:

36. Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). *British Medical Journal* 1992;305:875-877.

37. Cook J, Richardson J, Street A. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Economics* 1994;3:157-168.

Table 1. Socio-economic and clinical characteristics of women in the two valuation sample

Characteristics	Ex-ante HYE sample (n=63)	Health state valuation sample (n=60)	P value
Socio-demographic			
Mean (SE) age (years)	41.00 (0.92)	41.09 (0.77)	0.94(*)
Number (%) who have experienced serious illness:			
- themselves	15(24)	18(33)	0.28(#)
- in their family	41(65)	33(63)	0.86(#)
- in caring for others	20(32)	14(30)	0.83(#)
Number (%) smoking:			
- currently	17(27)	19(32)	0.47(#)
- formally	15(24)	18(30)	
- never	31(49)	23(38)	
Number (%) who have worked in health or social services	17(27)	22(37)	0.22(#)
Number (%) in employed	15(24)	33(55)	0.25(#)
Number (%) leaving school at minimum leaving age	31(49)	21(35)	0.16(#)
Number (%) with degree or equivalent professional qualification	8(13)	10(17)	0.53(#)
Clinical			
Median (range) duration of menorrhagia (months)	12(2-420)	24(3-360)	0.03(*)
Median (range) days per month bleeding	8(3-31)	8(4-20)	0.22(*)
Median (range) days per month with heavy flow	4(1-25)	4(2-14)	0.49(*)
Number (%) passing clots	54(92)	51(86)	0.38(#)
Number (%) with flooding episodes	53(91)	58(97)	0.27(**)
Maximum number of pads/tampons on heaviest day of period (numbers (%)):			
- 1-9	17(29)	18(31)	0.84(#)
- More than 9	41(71)	40(69)	
Median (range) days lost from work due to menstrual problems over the last year for those in work(##)	4(0-36)	2(0-48)	0.87(*)

* Wilcoxon rank-sum test
\# Chi Square test
** Fisher's exact test
n=35 (Swindon) and n=37 (Bristol)

Table 2. Health states values given by women using the time trade-off (n=60)

Health state	Mean (95% CI)	Median (Range)
Menhorragia	0.50 (0.42-0.58)	0.55 (0-0.95)
Pre-menopausal following recovery from successful TCRE	0.73 (0.65-0.81)	0.90 (0-1)
Pre-menopausal following recovery from AH	0.86 (0.80-0.92)	0.95 (0.05-1)
Convalescence following TCRE	0.76 (0.68-0.84)	0.85 (0-1)
Convalescence following AH	0.74 (0.64-0.84)	0.95 (0-1)

Table 3. Standard QALYs and *ex-ante* HYEs based on patient-specific data

Treatment	Mean (95% CI)	Median (Range)
Ex-ante* HYEs		
AH	34.84 (31.96-37.72)	37.5 (0-60)
TCRE	32.62 (29.52-35.72)	34.0 (0-60)
Standard QALYs		
6% discount rate		
AH	15.167 (14.792-15.543)	15.797 (9.795-16.197)
TCRE	14.380 (13.872-14.888)	15.409 (9.615-16.174)
2% discount rate		
AH	29.16 (28.94-29.38)	29.90 (22.76-30.37)
TCRE	28.23 (27.93-28.53)	29.45 (22.58-30.35)
0% discount rate		
AH	44.63 (44.38-44.88)	45.45 (37.62-45.95)
TCRE	43.62 (43.30-43.94)	44.95 (37.43-45.93)

* Mean years of future life used in the TTO was 44.76

Table 4. Standard QALYs and ex-ante HYEs for AH and TCRE: analysis of consistency based on the level of importance women attached to attributes in the questionnaire

Characteristic groups	N	AH>TCRE	TCRE>AH	AH=TCRE	Overall Inconsistency
Standard QALYs					**29/59 (49%)**
Hysterectomy	20	18 (90%)	2 (10%)		
TCRE	39	27 (69%)	8 (31%)		
Ex-ante HYEs					**17/62 (27%)**
Hysterectomy	19	12 (63%)	3 (16%)	4 (21.1%)	
TCRE	43	14 (33%)	12 (28%)	17 (40%)	

* One person in each valuation sample could not be allocated to a unique characteristic group as they had equal scores for both treatments

Table 5. Standard QALYs and *ex-ante* HYEs for AH and TCRE: analysis of consistency using women's choice of treatment based on the description provided in the questionnaire

Characteristics groups	N	AH>TCRE	TCRE>AH	AH=TCRE	Overall Inconsistency
Standard QALYs					**15/59 (25%)**
Hysterectomy	29	26 (87%)	3 (10%)		
TCRE	21	12 (62%)	9 (38%)		
Could not choose or would choose neither	9	8 (89%)	1 (11%)		
Ex-ante HYEs					**11/58 (19%)**
Hysterectomy	24	17 (71%)	2 (8%)	5 (21%)	
TCRE	24	6 (25%)	7 (29%)	11 (46%)	
Could not choose or would choose neither	10	3 (30%)	2 (20%)	5 (50%)	

Table 6. Standard QALYs and ex-ante HYEs according to women's stated positive and negative preferences for hysterectomy and 'other' treatments

Stated Preferences	N	Mean (SE)		Mean difference(95%CI)
		AH	TCRE	
Standard QALYs				
Positive stated preferences				
Hysterectomy	12	15.34(0.51)	14.03(0.69)	1.30(-0.20;2.64)
Other treatments	8	15.38(0.29)	13.82(0.71)	1.56(0.09;3.02)
No preferences	40	15.07(0.23)	14.59(0.29)	0.48(0.01;0.95)
Negative stated preferences				
Hysterectomy	9	14.71(0.48)	14.77(0.56)	-0.061(-0.56;0.44)
Other treatments	14	15.07(0.60)	13.66(0.67)	1.41(0.31-2.51)
No preferences	37	15.31(0.17)	14.55(0.29)	0.76(0.19;1.32)
Ex-ante HYEs				
Positive stated preferences				
Hysterectomy	9	35.89(2.82)	27.06(4.06)	8.83(-1.09;18.70)
Other treatments	6	32.33(4.67)	33.67(5.16)	-1.34(-14.20;16.80)
No preferences	34	33.51(2.26)	33.99(2.09)	-0.48(-5.66;6.62)
Negative stated preferences				
Hysterectomy	9	22.67(4.60)	23.33(4.49)	-0.66(-13.00;14.30)
Other treatments	6	37.42(4.29)	32.17(4.19)	5.25(-8.11;18.60)
No preferences	33	37.39(1.38)	33.50(2.11)	3.89(-1.15;8.93)

Appendix I Descriptive scenarios used in the *ex ante* HYE valuation study

TCRE

She suffers from heavy and painful periods and is about to have surgery for the condition. This will involve:

- a small risk of death of 1 in 1000;
- a stay in hospital of 1 day;
- an interval of about 1 week before she resumes her daily activities, during which she will experience some discomfort and sometimes feel tired;
- an interval of about 2 weeks before she can return to work;
- an interval of about 3 weeks before she can resume her sex life;
- the operation does not leave a scar.

Once she has recovered from the operation, she experiences the following results from surgery after about 4 months:

- she still has periods but they are much lighter since her operation;
- she still has some pain with her periods;
- she has no limitation on her social activities or daily activities such as work;
- she occasionally feels moody, irritable or depressed;
- she still has her womb, but it is unlikely that she will become pregnant;
- she is able to enjoy her sex life.

Two years after surgery she is likely to be happy with the results of her treatment and be in good health. However, during this time she would have faced the following risks:

- a 12% chance that she would have had the same operation again because of her menstrual problems returning;
- a 16% chance that she would have had another type of surgery because of her menstrual problems returning, involving 6 days in hospital and 4 weeks away from her usual activities.

Within 5 years after surgery any bleeding she still has ceases due to the start of her menopause.

She lives in good health for the remaining 35 years of her life.

AH

She suffers from heavy and painful periods in the same way as the person in the first description that you read, and is about to have surgery for the condition. This will involve:

- a small risk of death of 1 in 1000;
- a stay in hospital of about 6 days;
- an interval of about 4 weeks before she resumes her daily activities during which she will feel tired, need occasional pain killers and be unable to lift objects or walk very far;
- an interval of about 11 weeks before she can return to work;
- an interval of about 6 weeks before she can resume her sex life;
- she no longer has a womb, so she is unable to bear children;
- she will be left with a faint scar on her abdomen.

Once she has recovered from the operation, she experiences the following results from surgery after about 4 months:

- she no longer has periods or experiences pain;
- she has no limitation on her social activities or daily activities such as work;
- she occasionally feels moody, irritable or depressed;
- she is able to enjoy her sex life.

Two years after surgery she is happy with the results of her treatment and in good health.

Her menopause starts within 5 years after surgery and she lives in good

Appendix II List of treatment attributes in the questionnaire. Women were asked to rate the importance of each on a 4-point scale from 'very important' to 'not important'

A. I want to stay in hospital for as short a period as possible
B. I want treatment which will not remove my womb
C. I want treatment which will remove my womb
D. I want treatment which will not leave a scar on my abdomen
E. I want a treatment which causes the least pain and discomfort during my convalescence
F. I want a treatment to put a stop to my periods for good
G. I want treatment which will reduce my periods but not stop them for good
H. I don't want to have to worry about contraception after treatment
I. I want to resume my sex life as soon as possible after treatment
J. I want to be able to get back to work or carrying out my usual activities as soon as possible after my treatment