International Journal of Technology Assessment in Health Care, **18:3** (2002), 467–484.

HEALTHY-YEAR EQUIVALENTS IN MAJOR JOINT REPLACEMENT

Can Patients Provide Meaningful Responses?

Hilary A. Llewellyn-Thomas
Dartmouth Medical School

Rena Arshinoff
Mary Bell
Jack Ivan Williams
C. David Naylor
Sunnybrook and Women's College Health Sciences Centre

Abstract

Objectives: Healthy-years equivalents (HYEs) have been proposed as an evaluative measure with advantages over quality-adjusted life-years (QALYs). The main purpose was to assess the feasibility of eliciting HYEs from patients who have undergone major joint replacement; a secondary objective was to examine relationships with postsurgical health status.
Methods: Pre- and postsurgical reports of perceived comorbidity and current arthritic burden were obtained from 194 patients, using a comorbidity checklist, summary scores from the Western Ontario/McMaster Osteoarthritis Questionnaire (WOMAC), summary scores derived from six Likert scales, and holistic utility scores for the same attributes. After surgery, HYEs for the full across-time health profile were also elicited.
Results: All measures of arthritic burden were sensitive to pre/postsurgical changes ($p = .0001$), and comorbidity scores were stable. Two HYE subgroups emerged. An HYE-invariant subgroup ascribed full HYEs to their profiles, while reporting higher Likert ($t = 2.1309$; $p = .0344$) and utility ($s = 4.1504$; $p = .0001$) scores for their postsurgical health state. An HYE-variant subgroup reported HYEs that were weakly but significantly ($p < .009$) correlated with Likert ($r = .30$), utility ($r_s = .25$), and comorbidity ($r = -.26$) scores for their postsurgical state.
Conclusions: Our results indicate that patients can understand the HYE assessment procedures and provide interpretable responses. However, a significant proportion reports invariant HYEs that could inflate estimates of the overall mean HYE. Further exploration of the HYEs reported by different clinical and attitudinal populations is needed before widespread adoption of this measure.

This paper reports one aspect of the results obtained in a large research program. The Total Joint Replacement Research Group would like to acknowledge the support of the following orthopedic surgeons, who introduced the research program to the study participants: Chris Arciszewski, Robert Bourne, Hugh Cameron, Denis Desjardins, Jeffrey French, John Gordon, Allen Gross, Henry Hamilton, James Israel, James McAulay, Pram Mehta, John Porter, Michael Simurda, Francis Smith, Peter Southcott, John Stewart, and George Wong. In addition, the research group would like to acknowledge the careful work of the following interviewers, who worked with the study participants to collect the data used in this research program: Susan Berry, Joyce Brown, Susan Chamut, Jennifer Cross, Marilyn Braney, Aileen Davis, Kourosh Dinyari, Rose Gaudet, Suzanne Lacasse, Audrey Hunt, Joan Hutchings, Dana Meretsky, Rita Morriss, Karen Sullivan, and Elizabeth Whitton. Finally, we sincerely appreciate the contributions of the patients who participated in this study.

This study was supported by the Ontario Ministry of Health. Endorsement by the sponsoring agency is not implied.

Keywords: Quality-adjusted life-years, Arthritis, Arthroplasty

A MEASUREMENT PROBLEM

In cost-effectiveness analysis, the need for an outcome measure that simultaneously quantifies both survival time and health-related quality of life is a major measurement issue. Quality adjusted life-years (QALYs) have been advocated as a solution to this measurement problem (4;30;31;34;35).

The basic premise underlying the concept of the QALY is that it is appropriate to weight life years by a measure of the quality of life experienced during that time. It has been argued that not only should these weights be provided by the individuals who will receive the health services subject to QALY analyses, but also their attitudes toward time and risk should be included in the assessment process (20;21). Thus, a full representation of the relevant preferences would incorporate individuals' time discount, their risk attitude regarding gain and loss in survival time, and their risk attitude regarding gain and loss in quality of life.

The extent to which this full representation can be approximated depends on: a) the time frame surrounding the assessment process; b) the particular strategy used to elicit the "quality" weight for the relevant states of health; and c) the stability of the profile of health states across time. The first two sets of issues are outlined in Figure 1.

	TIME FRAME	
ELICITATION STRATEGY	**Variable Time Frame** (1 + 1…25 yrs)	**Predetermined Time Frame** (Full 25 yrs)
Value Assessment	TTO value for 1 yr × 25 × time discount	TTO value for full 25-year profile
Time Discount?	attributed	incorporated
Risk Attitute re: Survival?	missing	missing
Risk Attitude re: Q of Life?	missing	missing
Utility Assessment	SG utility for 1 yr × 25 × time discount	SG utility for full 25-yr profile
Time Discount?	attributed	incorporated
Risk Attitute re: Survival?	missing	missing
Risk Attitude re: Q of Life?	incorporated	incorporated

Figure 1. Different approaches to assessing QALYs: Representing individual preferences.

Assessment Time Frame

The time frame could be either variable or predetermined, requiring incremental or holistic approaches, respectively, to QALYs estimation. The incremental strategy involves first assessing the quality weight for 1 year in a particular state of health and then multiplying the number of years of interest by this quality weight. The holistic strategy involves eliciting the overall quality weight for the full profile of years in the particular health state.

At first glance, the incremental strategy seems to have the advantage of flexibility, in that it requires only one preference assessment, and QALYs for various lengths of time can subsequently be inferred. Accordingly, it would be most useful in QALY analyses of clinical situations in which interventions have different effects on the extension of life expectancy. Since separate assessments would have to be carried out for full profiles of different durations, the holistic strategy would appear to be confined to the evaluation of clinical situations in which interventions affect quality of life rather than its duration (QALY analysis). However, in order to draw inferences from the first strategy, one must make assumptions about the individuals' time discounts, whereas these are incorporated in assessments obtained with the second strategy.

Elicitation Strategies

Besides these time frame considerations, whether or not the quality weights are elicited by methods incorporating uncertainty affects the extent to which fully representative QALYs can be derived. Often, weights generated by risk-free methods such as the time trade-off are expressed as value scores, and those derived with risk-based techniques such as the standard gamble are referred to as utility scores (28;32).

When used to evaluate a full profile of several years, the time trade-off technique obtains in a single step the time-discounted value-weighted QALYs for that profile. Suppose, for example, 25 years in renal dialysis are considered equivalent to 20 years in good health, or 20 QALYs. If one is willing to make the arguable assumption of a constant proportional trade-off between time and quality (18), this result could in turn be used to derive a value weight for 1 year of dialysis (i.e., $20/25 = .80$), and value-weighted QALYs for time periods of other durations could be subsequently computed. However, because the process of eliciting the time trade-off value does not involve considerations of uncertainty, the individual's attitudes toward risks related to survival time and quality of life are not incorporated in the QALYs obtained using either time frame.

Because the standard gamble involves considerations of gain and loss in health status, the utility-weighted QALYs generated by this method do incorporate the individual's attitude toward risks related to quality of life, but still do not include risk attitude toward survival time. Accordingly, Mehrez and Gafni (20) maintain that the utility-based version of the QALY has to be based on restrictive assumptions about the form of the individual's utility function.

Stability of Health Profiles

Furthermore, both the risk-free and risk-based methods for obtaining quality weights greatly simplify the clinical situation, in that they imply that an individual remains in the same health state over the full time period under consideration. In actuality, an individual often experiences a series of different health states over time. For example, a health profile may involve a period of time in a reasonably good health state that then becomes problematic and requires an intervention that, in turn, makes the individual temporarily worse. This may be followed by a recovery time, a rehabilitation period, and finally a subsequent postintervention period in which the individual feels better than during the immediate preintervention period.

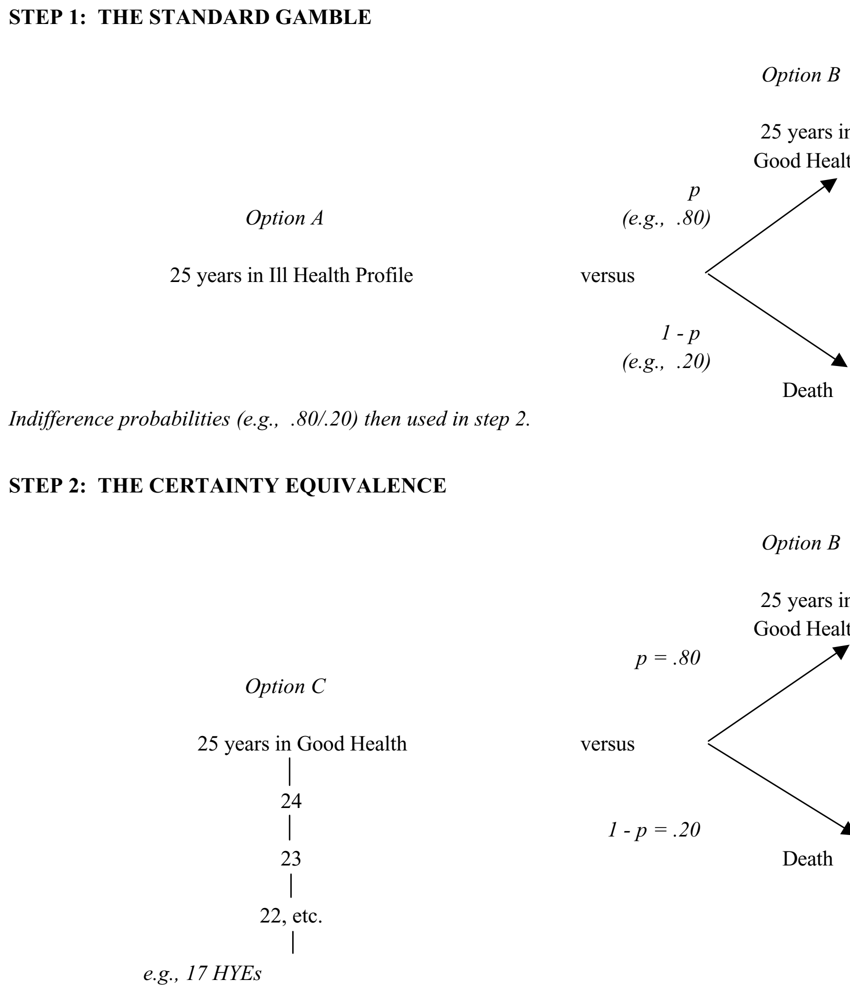

Figure 2. Assessing HYEs.

Thus, in its fullest form, the problem is to identify a measurement approach that permits an overall evaluation of a profile of fluctuating states, while simultaneously incorporating individuals' time discounts, risk attitudes toward health status, and risk attitudes toward survival time. Mehrez and Gafni (20;21) propose that HYEs solve these dilemmas.

ASSESSING HYEs

Before outlining the objectives of this study, it is necessary to describe the procedures involved in assessing an individual's HYEs for a profile of health states that he or she has actually experienced. To begin, the respondent is asked to think back over this profile, then to consider re-experiencing it and continuing into the future in his/her current health state. Then the assessor proceeds with two steps, as illustrated in Figure 2.

In the first step, the standard gamble technique—which traditionally is used to elicit utilities for single states—is used to obtain a utility for this full, multiple-state health profile. The respondent is asked to consider a hypothetical choice between, for example, 25 years in this health profile and a gamble. The gamble has two possible outcomes. The positive outcome is a guarantee of 25 years in good health followed by death (arbitrarily assigned a utility of 1.0). The negative outcome is immediate death (arbitrarily assigned a utility of 0.0).

The probabilities in the gamble are systematically altered until the respondent cannot choose (i.e., is indifferent) between the certainty of 25 years in his/her health profile and the gamble. Suppose, for illustrative purposes, the respondent is indifferent between the 25-year certainty and the gamble, when $p = .80$ for 25 years in good health (and, by corollary, the probability of death is $1 - p = .20$). This indifference probability is then used in the second step, which involves the certainty equivalence technique (19;29).

In the certainty equivalence technique (see Figure 2, step 2), the problem structure is initially identical, in that the respondent is presented with a choice between a certainty—in this case, of 25 years in *good health*—and a gamble with the probabilities that were originally identified in step 1 (here .80/.20). The logical response would be to choose the guarantee of 25 years in good health for certain. At this point in the certainty equivalence technique (unlike the procedure followed in the standard gamble), the probabilities are held constant, and the time in the certainty is varied; i.e., the number of years in good health are systematically altered (for example, by lowering to 24, then 23, etc.) until the person has difficulty (say, at 17 years) with choosing between good health for certain and the .80/.20 gamble.

In our example, 17 years is considered to be the respondent's HYEs for his/her ill health profile. This conclusion is based on the following logic. Since the gamble with the probabilities of .80/.20 (option B) was originally, in step 1, considered by the respondent to be equivalent to 25 years in the health profile (option A), and since in step 2 option B is considered equivalent to 17 years in good health (option C), therefore, by substitution, 25 years in the ill health profile is equivalent to 17 HYEs (option C = option A).

OBJECTIVES

Johannesson et al. (13) have challenged Mehrez and Gafni's conceptualization of HYEs, arguing that the certainty equivalence in step 2 removes the risk attitude introduced in step 1, so that the end result is essentially a risk-free time trade-off value for the full profile. Mehrez and Gafni's rebuttal in effect repeats the argument illustrated in Figure 2, maintaining that individuals incorporate their risk attitude regarding quality of life into their responses to step 1, and their risk attitude toward survival time into step 2 (22). There has been an avid HYEs versus QALYs debate in the preference measurement literature (5;6;7;11;12;13;17). However, this debate may largely be fueled by mutually unrecognized differences in the conceptualization and use of the term *risk*; often these opposing authors, when they develop their arguments, are not explicit about whether they are referring to risk attitudes about losses in quality of life or about losses in survival time. This impression is reinforced by Ried's very helpful overview of the debate (26). Ried concludes that, with its focus on health status sequences, the HYE is a particularly fruitful methodologic route for further exploration of patients' preferences. Dolan (8;9) also does not dismiss the potential usefulness of the HYE method. The resolution of the conceptual controversies is beyond the scope of this paper. Instead, given the minimal empirical work with this potentially useful approach, our objective was much more modest: we were primarily interested in the feasibility of carrying out this measurement method with patients who were experiencing actual health profiles in real time.

We assumed that an empirical exploration of HYEs would involve eliciting responses from individuals who were chronically ill, had undergone an intervention specific to their condition, and were able to reflect upon their entire health profile. A longitudinal study of patients undergoing major joint replacement would be an appropriate clinical context in which to carry out such exploratory work. Furthermore, this clinical group could provide self-reports about their health status (expressed in general comorbid or arthritis-specific terms) at the time HYEs are elicited. Thus, the main objective of this study was to determine whether it was possible to obtain HYEs from patients who have undergone major joint replacement. A secondary objective was to explore possible relationships between those HYEs and covariates such as self-reports about postsurgical comorbid and arthritis-specific health status at the time of HYE elicitation.

METHODS

Recruitment of Participants

This study formed one component of a large Ontario-wide patient survey addressing a number of health services research issues in total joint replacement (15;36). Nineteen collaborating orthopedic surgeons in eight Ontario cities agreed to contact their patients who were awaiting primary hip or knee arthroplasty. Outlying areas of each city were included as part of the catchment area for the survey.

Eligibility criteria for entry into the full survey required that the patient either have a specific date for a primary total hip or total knee replacement or be on a waiting list for a booked date be able to speak, read, and understand English, and be over 18 years of age. All 471 apparently eligible patients received a letter from their own surgeon that introduced and outlined the study and encouraged participation. Each patient was then contacted, and formal informed consent to participate was sought. Of the 471 referred patients, 413 were actually eligible, and 322 of these could be contacted. Of the 322 eligible available patients, 47 refused participation; thus, with 275 participants at intake, the initial overall response rate for the full survey was 85.4%. Of these 275 participants, for the purposes of the analysis described here, if any of an individual's data points at the intake interview were missing (n = 17), his/her results were eliminated from the data set. At the time of the postsurgical interview (see below), 35 were either lost to follow-up or had incomplete data, and 29 had not yet had their surgery, thereby rendering a total of 194 eligible patients for whom complete sets of pre- and postsurgical interview data were obtained.

Data Collection

General Approach. Each study participant was interviewed on two occasions by 1 of 15 trained interviewers. All interviews were conducted in the patients' homes at their convenience. In order to promote a standardized interview process throughout the province, the interviewers were formally taught the interviewing techniques at 2-day study training sessions prior to each interview. At the beginning of both interviews, the study's purpose and procedures were described, the voluntary nature of participation was emphasized, the confidentiality of data was assured, and informed consent to proceed with the interview was obtained.

Initial Interview: Assessing Covariates Before Surgery. At the time of the first interview, participants were either waiting or had actually been booked for their surgery. The purpose of this interview was to obtain demographic data as well as baseline reports for our postulated covariates.

PERCEIVED COMORBIDITY

One possible covariate was the respondents' perceived comorbidity. The respondent considered a comorbidity checklist of 18 different illnesses/conditions other than arthritis (for example, cardiac disease, metabolic disorders, cancer, and other systemic conditions) that could be associated with functional impairment. For each item, the patient indicated if he/she was free from the problem ("no problem" = 0), experienced the problem but it did

https://doi.org/10.1017/S0266462302000326 Published online by Cambridge University Press

not limit activities (“yes, no limitation” = 1), or experienced it with activity limitation (“yes, limitation” = 2). These scores were summed and normalized to obtain a total comorbidity score that could range from 0 (no comorbidity) to 100 (high comorbidity). Note that this index was designed to obtain an objectively weighted score for the patient’s perceived comorbidity, rather than to serve as an objective measure of actual comorbidity, since the patient’s awareness of comorbidity was more germane to the problem of exploring the determinants of the HYEs.

ARTHRITIS-SPECIFIC HEALTH STATUS

The other postulated covariate was self-reported arthritis-specific health status. We could have elicited baseline data for this covariate using a single set of arthritis-specific attributes, coupled with an entirely objectively-weighted scoring scheme. However, the ability to detect apparent relationships with HYES, which are inherently subjectively weighted, could be affected by the extent to which the covariate’s response scale allowed subjective weights to come into play. Therefore, we used three approaches which incorporated comparable arthritis-specific attributes but varied in terms of their latitude for subjectivity in scoring:

1. *The WOMAC*. First, the visual analog version of the Western Ontario/McMaster Osteoarthritis Questionnaire (WOMAC) was used to obtain a summary score describing the respondent’s arthritic state relative to his/her major joint problems. The WOMAC was developed by Bellamy and colleagues (1;2;3) to assess clinically important patient-relevant outcomes. Evidence for its validity, reliability, and sensitivity appears in reports evaluating hip or knee arthroplasty, antirheumatic drug therapy, and pain in patients with arthritis. This instrument assesses three domains: pain (5 items), stiffness (2 items), and function (17 items). For each item, the subject marked a 10-centimeter linear scale indicating his/her current status relative to the scale’s anchors. The left anchor indicated no pain, stiffness, or dysfunction while the right side indicated extreme levels. After the interview, for each scale, the distance from the left-hand anchor to the patient’s mark was recorded in millimeters. Lower scores reflected lower levels of pain, stiffness, or dysfunction, while higher scores represented more intense difficulties in these domains. A total WOMAC score was obtained for each patient by calculating a mean ranging between 0 to 100 for each of the three assessment domains, then adding these transformed means to yield a maximum possible score of 300. Thus, although using visual analog scales allows some subjectivity into the raw scores, the steps required to aggregate across the three domains implies that each is equally weighted and controls for the number of items in each domain. Then, for the purposes of this analysis, the totals were reverse-scored and transformed to range from 0 to 100, with higher values representing better levels of arthritis-related health status.
2. *Likert scales*. Next, respondents reported whether they currently experienced no, mild, moderate, or severe distress in each of six arthritis-specific attributes: pain, stiffness, difficulty doing home/work activities, difficulty doing self-care activities, restrictions in leisure activities, and emotional distress due to arthritis. To obtain Likert evaluative scores, these self-reports were later scored as follows: no = 4; mild = 3; moderate = 2; and severe = 1. After the interview, an objective summary score was obtained by adding the six ratings, yielding a scale that ranged from 6 to 24, with higher scores representing higher levels of arthritis-related quality of life. These summed scores were transformed to range from 1 to 100.
3. *Utilities*. Finally, to obtain the most subjectively weighted evaluative scores for current arthritis-specific health status, we used the standard gamble as it is applied to single states. To do this, the six self-reports that had been obtained when eliciting the Likert evaluations were used to create, on the spot, an individualized descriptive ‘scenario’ for the patient’s current presurgical health state. Then the respondent considered a hypothetical choice between continued life in this scenario and a gamble. The positive gamble outcome was a health state with no distress in the six dimensions (the “arthritis-free” health state, arbitrarily assigned a value of 100). The negative outcome was a health state with severe distress in all six dimensions (the “worst-arthritis” health state, arbitrarily assigned a value of 0).

A side note is necessary here. Generally, good health and death are used as the outcomes in the single-state standard gamble, when the main research objective is to elicit utilities for the purpose of making across-disease comparisons. In our case, the need to make across-disease comparisons was not the motivation for collecting current health state utilities in this project—recall that we merely wished to obtain covariate utilities for the arthritis-specific current health state. It was necessary to do this in a way that avoided the confounding effects of comorbidity that are introduced when death is used in gambles attempting to reveal disease-specific utilities (23). Furthermore, to maintain internal consistency across all three measures for this covariate, it was also necessary to do this in a way that was conceptually congruent with the WOMAC and Likert scales, which involved assessment of similar constituent attributes without reference to the extremes of good health and death (14;16). Therefore, given these multiple considerations, we used the 'best-arthritis' and 'worst-arthritis' states as our single-state gamble outcomes.

The probabilities were shifted until the respondent could not choose between the certainty of continued life in his/her presurgical health state and the gamble. The expected value of the gamble at this point is, by substitution, the utility for the arthritis-specific health state (28), and, in this analysis, ranged from 0 to 100.

Second Interview: Assessing Covariates and HYEs After Surgery. All participants were reinterviewed 1 year later, when the participants were at various postsurgical time points.

ASSESSING THE COVARIATES

At the beginning of the session, we repeated the comorbidity, WOMAC, Likert, and utility assessments that were used in the first interview. These were our key covariate data—that is, respondents' self-reported comorbid and arthritis-specific health status at the time of HYE elicitation.

ASSESSING THE HYEs

Next, we carried out the data collection step of primary interest—that is, the elicitation of each patient's HYEs for his/her overall 'presurgical + postsurgical' health profile. This began with constructing a visual timeline to review: a) each respondent's experience with arthritis up until his or her initial interview; b) his/her individualized presurgical scenario that had been constructed in the first interview; c) the recent experience of undergoing and recovering from surgery; and d) the individualized postsurgical scenario that had just been constructed in this second interview. The respondent was asked to imagine re-experiencing the past year (including b, c, and d) as well as continuing into the future for 24 years in his/her current postsurgical health state (for an overall health profile of 25 years). Once the respondent clearly understood that he/she was being asked to evaluate this 25-year multiple-state profile, the interviewer proceeded with the two-step HYE assessment outlined in the introduction.

In the first step, the respondent considered a hypothetical choice between his/her 25-year health profile and a gamble. Here, the gamble took the form advocated for HYE assessment, in that the positive outcome was a guarantee of 25 years in good health followed by death and the negative outcome was immediate death. The probabilities in the gamble were varied until the respondent's indifference point was identified.

This indifference probability was then used in the second step, involving the certainty equivalence technique. The respondent was asked to choose between a certainty of 25 years in good health and a gamble with the indifference point probabilities he/she had identified in the first step. The respondent logically chose the guarantee of 25 years in good health for certain. Then the gamble probabilities were held constant, and the time in the certainty was lowered until the person had difficulty choosing. Then, according to the logic outlined in the introduction, the years in good health at this point represented the respondent's HYEs for his/her ill health profile.

One could argue that, ideally, our exploration of the HYE method should involve the real-time construction of a longer series of health states over an extensive time period into the future. This would come closer to assessing how well the HYE captures attitudes toward actual fluctuating health profiles, which is putatively one of its advantages. However, the conduct of such an extensive longitudinal design in real time would be difficult and expensive. Our project was driven by a more immediate interest in assessing the feasibility of obtaining responses to the two steps in the HYE elicitation task. These two steps reflect the other putative advantage of the HYE—that is, its ability to differentially tap into individuals' risk attitudes regarding quality of life and survival time.

Data Analysis Plan

Descriptive statistics were used to summarize the data distributions of the respondents' sociodemographic and clinical characteristics, their HYEs, their scores for the comorbidity covariate, and their WOMAC/Likert/utility scores for the arthritis-specific health status covariate. To evaluate the sensitivity of the covariate measures, we assessed across-time differences using the paired t test (comorbidity, WOMAC, and Likert scores) and the Wilcoxon signed rank test (utility scores). Then Pearson and Spearman correlation coefficients were computed for the relationships between HYEs and the covariates.

RESULTS

Patient Characteristics

Table 1 highlights selected sociodemographic and clinical characteristics of the full sample of participants. There were more women than men, and the mean age was 66 years. Most of the participants were retired, married, and living with their spouse. Osteoarthritis was the most common reason for the planned joint replacement.

Overall Performance of the Covariate Measures

Table 2 describes the distributions of the pre- and postsurgery comorbidity, WOMAC, Likert, and utility scores for the full sample. We compared these pre- and postsurgery scores in order to investigate the stability of the comorbidity measure as well as the sensitivity of the other health status measures to anticipated across-time change. As expected, the mean comorbidity score did not change as the patients moved from pre- to postsurgery, and there were statistically significant across-time increases in mean scores on the disease-specific WOMAC, Likert, and utility scales, indicating improvements in the levels of distress and dysfunction associated with arthritis in the affected joint. Accordingly, we concluded that our covariate measures were behaving reasonably in this application, and we could proceed to explore their relationships with the HYEs themselves.

Distributions of Observed HYEs

The first column in Table 3 describes the distribution of HYEs for the full sample. Overall, the HYEs ranged from 1 to 25, with a mean of almost 20 HYEs. At this point, we could have proceeded to examine the overall relationships between the HYEs and each of the covariate health status measures.

Emergence of Two HYE Subgroups. We noted, however, that a large proportion of respondents ($n = 83$; 43%) chose not to gamble in step 1 of the HYE assessment task (Figure 1). These respondents were reporting an indifference probability of 1.0 in step 1, and

Table 1. Selected Sociodemographic and Clinical Characteristics of 194 Patients Who Underwent Total Joint Replacement Surgery

Characteristic	n	%
Male	80	41.2
Female	114	58.8
Married	125	64.4
Single	18	9.3
Divorced	9	4.6
Widowed	42	21.6
Living alone	39	20.1
With spouse	120	61.9
With other family	26	13.4
Other	9	4.6
Mean Age: 65.9 yrs		
Age < 50 yrs	17	8.8
50-59	26	13.4
60-69	63	32.5
70-79	76	39.2
80+	12	6.2
Paid employment	44	22.8
Retired	79	40.9
Keeping house	53	27.5
School	2	1.0
Other	15	7.8
Osteoarthritis	145	74.7
Rheumatoid arthritis	18	9.3
Other	26	13.4
No answer	5	2.6

Table 2. Pre- and Postsurgery Health Status Reported by Full Sample: Distributions and Tests for Across-time Differences in Co-morbidity (*t* Test), WOMAC (*t* Test), Likert (*t* Test), and Utility (Wilcoxon) Scores (n = 194)

	Co-morbidity		WOMAC		Likert		Utility	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Mean	17.12	17.98	43.20	73.30	36.40	76.58	65.91	92.98
Standard Deviation	7.66	8.03	17.20	19.30	19.20	16.60	26.23	17.26
Minimum	2.94	5.88	5.69	11.01	0.00	6.00	0.00	1.00
Maximum	58.82	52.94	91.90	100	94.00	100	100	100
Statistics	—		$t = 20.99$		$t = 23.67$		$s = 7{,}030$	
p	NS		.0001		.0001		.0001	

Table 3. HYEs Reported by Full Sample and by HYE-variant Subgroup

	Full sample (n = 194)	HYE-variant subgroup (n = 111)
Mean	19.90	16.08
Median	22.00	15.0
Standard Deviation	5.83	5.02
Minimum	1	1
Maximum	25	24

Table 4. HYE-variant Subgroup's Pre- and Post-surgery Health Status Scores: Across-time Differences in Comorbidity (*t* Test), WOMAC (*t* Test), Likert (*t* Test), and Utility (Wilcoxon) (n = 111)

	Presurgery	Postsurgery	
Comorbidity			
Mean	17.01	17.99	
Standard Deviation	7.55	7.32	$t = 1.62$
Minimum	5.88	5.88	$p = .1080$
Maximum	58.82	41.18	
WOMAC			
Mean	43.14	71.47	
Standard Deviation	17.53	20.16	$t = 14.86$
Minimum	5.69	11.37	$p = .0001$
Maximum	86.99	100.00	
Likert			
Mean	35.79	73.77	
Standard Deviation	19.0	22.73	$t = 16.53$
Minimum	6	6	$p = .0001$
Maximum	100	100	
Utility			
Mean	64.62	88.59	
Standard Deviation	25.21	21.68	$s = 2{,}151$
Minimum	0	1	$p = .0001$
Maximum	100	100	

therefore each was indicating that their health profile was equivalent to 25 HYEs. When this subgroup, with their invariant HYEs, was removed from the full sample, the mean HYEs dropped from approximately 20 years to 16 years, as seen in the second column in Table 3. Thus, the respondents who reported variant HYEs (n = 111; 57%) were indicating that their health profile of 25 years (which included their experience with arthritis, surgery, and recovery) was, on average, equal to 16 healthy years.

Performance of Covariate Measures by HYE Subgroup. We repeated our earlier assessment of the performance of the covariate measures for each of these newly identified HYE subgroups. Tables 4 and 5 present the mean across-time differences in pre- and postsurgical scores reported by those who reported variant and invariant HYEs, respectively. In both subgroups, the change in comorbidity scores was nonsignificant, while statistically significant change scores were observed for the WOMAC, Likert, and utility scales. This indicated that, for both HYE subgroups, the covariate measures continued to demonstrate the same pattern of sensitivity and stability that had been noted for the full group.

HYE Subgroup Comparisons. These two HYE subgroups did not differ on any of the sociodemographic and clinical characteristics outlined in Table 1. We then determined whether the subgroups differed in covariate health status at either the presurgery or postsurgery times, or in time since surgery at the second interview. Table 6 indicates that there were no statistically significant differences between the two subgroups before surgery on any of the variables. Table 7 indicates that mean scores for the comorbidity index, the WOMAC, and time since surgery also did not differ across subgroups after surgery. However, the two HYE subgroups reported significantly different Likert and utility scores after surgery, in that those who provided invariant 25-year HYEs also provided higher Likert and utility scores.

Table 5. HYE-invariant Subgroup's Pre- and Postsurgery Health Status Scores: Across-time Differences in Comorbidity (t Test), WOMAC (t test), Likert (t test), and Utility (Wilcoxon) (n = 83)

	Presurgery	Postsurgery	
Comorbidity			
Mean	17.26	17.97	
Standard Deviation	7.85	8.93	$t = 1.27$
Minimum	2.94	5.88	$p = .2078$
Maximum	41.18	52.94	
WOMAC			
Mean	43.26	75.68	
Standard Deviation	16.93	18.00	$t = 15.02$
Minimum	8.37	11.01	$p = .0001$
Maximum	91.86	99.24	
Likert			
Mean	37.08	80.36	
Standard Deviation	20.03	19.28	$t = 17.32$
Minimum	0	11	$p = .0001$
Maximum	94	100	
Utility			
Mean	67.63	98.86	
Standard Deviation	27.59	3.02	$s = 1416$
Minimum	0	85	$p = .0001$
Maximum	100	100	

Table 6. Subgroup Differences Before Surgery: Means and Test Results for Comorbidity, WOMAC, Likert, and Utility Scores

	Variant HYEs (n = 111)	Invariant HYEs (n = 83)	
Comorbidity	17.01	17.26	
WOMAC	43.14	43.26	All nonsignificant
Likert	35.79	37.08	
Utility	64.62	67.63	

Table 7. Differences Across Subgroups After Surgery: Means and Test Results for Comorbidity, WOMAC, Likert, and Utility Scores, and Number of Weeks Since Surgery

	Variant HYEs (n = 111)	Invariant HYEs (n = 83)	Statistic	p
Comorbidity	17.01	17.26	$t = 0.0213$	.9831
WOMAC	71.47	75.68	$t = 1.5045$	.1341
Likert	73.77	80.36	$t = 2.1309$	.0344
Utility	88.59	98.86	s = 4.1504	.0001
Weeks since surgery	41.64	42.81	$t = 0.6549$	.5133

The Relationship Between Covariate Health Status and HYEs

Recall that our secondary purpose was to assess the relationship between the HYEs scores and the covariate health status measures after surgery. We could proceed to examine these relationships only in the HYE-variant subgroup.

Table 8. HYEs-variant Subgroup: Correlation Coefficients for Relationships Between HYEs and Postsurgery Health Status Scores ($n = 111$)

	Coefficient	*p*
Comorbidity	$r = -.26$	.005
WOMAC	$r = 0.16$	.0961
Likert	$r = 0.30$	.001
Utility	$r_s = 0.25$	.009

Table 8 presents the correlation coefficients for the relationships between the HYEs and the health status scores in this sub group. Due to the skewed distribution of the utility scores, Spearman's correlation coefficient was used to test the relationship involving utilities, whereas the Pearson coefficient was used with the other covariate health status measures. HYEs were not related with the WOMAC scores but were significantly weakly correlated with the postsurgery comorbidity, Likert, and utility scores.

DISCUSSION

The Feasibility of the HYE Task

HYEs are advocated by Mehrez and Gafni as an aggregate measure of both the quantity and quality of life, which avoids restrictive assumptions about the individual's preferences regarding time and risk. This study was not designed to resolve the theoretical debates surrounding the concept of HYEs. Our primary objective was to assess whether patients undergoing major joint replacement surgery could actually engage with the HYE task and provide interpretable responses.

All of our respondents were able to do this, indicating that it is feasible to carry out the HYE assessment procedures with individuals who are chronically ill with a non–life-threatening disorder and who, across time, have experienced a changing series of health states. All respondents were able to consider the overall health profile they had experienced leading up to and recovering from their surgery, and to comprehend the multiple-state standard gamble and certainty equivalence techniques used to derive the HYEs for their profile.

As noted above, a full exploration of the feasibility of the HYE method would ideally involve the real-time assessment of a health profile, consisting of a series of more widely fluctuating health states unfolding over a more extensive time period. Still, the outcomes of the limited study reported here imply that raw HYEs can be elicited from individuals in real clinical settings, as an initial step in obtaining a distribution of HYEs in a given clinical population.

However, caution is indicated. Other investigators have also observed variant and invariant subgroups in utility distributions, although these reports have been in the context of single-state descriptions only (10;24;27). Taken together, these results indicate that in both single-state and profile assessments, individuals may differ widely in their attitudes toward risking losses in their current quality of life. Suppose an investigator, given a particular multiple-state profile, wishes to estimate a particular population parameter such as the mean HYEs for that profile, so that this mean can be used in subsequent policy decisions. If the presence of variant and invariant subgroups is ignored, the investigator runs the risk of under- or overestimating that mean, which could in turn have ethical as well as scientific consequences.

Are HYEs Affected by Concurrent Health Status?

Our secondary objective was to explore for possible relationships between HYEs and comorbidity and arthritis-specific health status at the time of HYE assessment. Our results raise several theoretical and methodologic issues about attempting to characterize possible determinants of reported HYEs.

The HYEs-Invariant Subgroup. Correlational analytic strategies were not appropriate with this subgroup, and we can offer only speculative comments about the possible causes of their relatively high HYEs. One possibility is that the members of this subgroup are characteristically "nongamblers" in all assessment tasks involving considerations of risk. However, these patients did engage with the gamble that was used to elicit disease-specific utilities for their presurgical health status, and provided responses that were comparable to those of the HYEs-variant subgroup (compare the disease-specific utilities reported in Table 5 with those in Table 4). It is possible that the experience of surgery somehow led these patients to become particularly risk-averse, and their higher postsurgical utilities and invariant HYEs merely reflect that transformation. However, their postsurgical Likert scores are also significantly higher, and considerations of risk were not involved in derivation of the Likert scores.

Therefore, perhaps the HYE-invariant responses emerge not because of differential attitudes towards risk *per se*, but because these individuals did not actually incorporate their long-term experience into their HYE reports. Instead, they may have been particularly affected by their strong sense of well-being after surgery. Recall that this subgroup did not differ on any of the demographic, clinical, or health status covariates at the presurgical interview, nor did they differ in time since surgery at the postsurgical interview. One might anticipate that these respondents would behave similarly after surgery. However, they reported a distinctly greater degree of improvement brought about by surgery, as implied by their significantly higher Likert and utility scores. If satisfaction with their postsurgical state washes out judgments of the relative undesirability of their presurgical condition, they would, as observed here, consider the gamble in step 1 not to be 'worthwhile' (which in turn translated into high HYEs values of 25 years for the full profile).

Accordingly, this subgroup may be manifesting a kind of judgment heuristic (33) in the form of a "recency" effect—that is, their overall evaluations of the series of changes in their health status were affected by the level of perceived well-being in the last phase of the experience. This phenomenon has been postulated by Redelmeier and Kahneman (25) to account for their observations that retrospective evaluations of a painful episode were higher than predicted by the evaluations collected during the episode.

The HYEs-Variant Subgroup. Within this subgroup, statistically significant correlations with the HYEs were weakly positive for the postsurgery Likert and utility scores, and weakly negative for the comorbidity scores, while HYEs did not correlate with the postsurgery WOMAC scores. There are several points for discussion here.

The first point involves the directions of the observed relationships. The inverse relationship with comorbidity may indicate that these respondents were at least considering the long-term aspects of their health profile when they reported their HYEs. The positive relationships could imply that HYEs are also somewhat affected by respondents' current health status at the time of HYE elicitation, whether current health status is evaluated using more objective (Likert) or subjective (utility) approaches. (However, we cannot determine if the recency effect postulated above is also operating—albeit to a lesser degree—in the HYE-variant subgroup. This is a worthwhile area for further investigation, because a recency effect would introduce a further cautionary note about generalizing even from an HYE-variant group to the larger population. The variant group's mean HYEs may be biased upward by those respondents who are manifesting the recency effect, and, consequently, invalid conclusions inadvertently might be drawn about the overall evaluation of the 'average' health profile experienced by individuals in that particular clinical situation.)

The second point is that these were not strong correlations. There are a number of possible accounts for this observation. Our measures may simply have been incapable of detecting actual stronger underlying relationships. However, a range of scores had been observed in this subgroup's distribution of HYEs (Table 3), and appropriate across-time patterns had been observed for the Likert and utility scores (Table 4); these results imply that the measures were not wholly insensitive to existing variance. The low correlation between HYEs and health status utilities—both of which involve gambles—may arise because death is considered in the former and not in the latter. However, a correlation of comparable magnitude was observed with the Likert method, which does not involve gambles at all. Finally, it is necessary to point out that there is no strong *a priori* reason for assuming that there *must* be a strong relationship between HYEs and current health status at the time HYEs are reported. It is possible that these individuals' attitudes toward their full multistate HYE health profile are relatively impervious to influences from the health state they happen to be experiencing at the time of HYE assessment. If this were so, then weak correlations would be observed, as reported here.

The third noteworthy point is that the WOMAC summary scores after surgery did not correlate with this subgroup's reported HYEs. As argued above, this lack of correlation did not seem to arise because of insensitivity of the measures. Furthermore, the different performance of the WOMAC scores relative to the Likert and utility scores does not appear to be due to inability to tap into respondent subjectivity. Although the WOMAC is admittedly less engaging than the standard gamble, our use of visual analog scales to assess perceived levels of distress should allow greater subjectivity to come into play with the WOMAC compared with the fixed-response format used in the Likert technique.

It seems more probable that the lack of correlation between this subgroup's HYE and WOMAC scores is due to an artifact introduced by our approach to summing and transforming the WOMAC raw scores. Although there were 17 items in the WOMAC's function domain, the raw scores were summarized in a single mean value for that domain. On the other hand, the Likert scoring system allowed three functional areas (household/work, self-care, and leisure) to contribute separately to the Likert summative score, and those three areas were also explicitly presented in the scenarios used in the subsequent utility assessment exercise as well as in the construction of the HYE profile itself. Thus, although the WOMAC linear analog scales may allow some subjectivity to come into play, their scoring system in effect "dampened down" the reported distress associated with functional disability, with a consequent lack of correlation with the HYEs.

In summary, our examination of relationships between HYEs and evaluative scores for concurrent health was exploratory only; it was not undertaken to test hypotheses about the validity of HYE assessment. Taken together, these discussion points merely highlight the conceptual and methodologic issues to consider in future attempts to determine whether current health status affects reported HYEs. This discussion also raises two larger questions: Is it important or worthwhile to try to characterize the determinants of HYEs, in the first place? If so, what determinants other than current health should be considered?

CONCLUSIONS

Policy Implications

These patients were able to engage with the HYEs assessment task, implying that it is operationally feasible. This in turn implies that it is possible to carry out HYE elicitation in other clinical groups, when the ultimate aim is to use these attitudinal data in subsequent decision analyses that will guide the development of health policies. However, as noted in earlier work involving single-state descriptions, there were different attitudinal subgroups underlying our observed overall distribution of HYEs ascribed to multistate health profiles. Investigators who wish to draw valid generalizations from HYE distributions should be alert to the possible hidden presence of these subgroups.

Theoretical and Methodologic Implications

One postulated determinant of these attitudinal subgroups is the degree to which different individuals are susceptible to a recency effect when providing HYE responses. If such an effect exists, it threatens the ability to elicit evaluations for profiles of health states that unfold over time, which is one of the purported advantages of the HYE. In order to systematically test hypotheses about such effects, investigators will need to devise ways to measure attitudes toward current health status at various points in time in the profile. In doing so, they should be alert to possible artifacts that can be introduced by the measurement strategies themselves. Attention should be paid to the multiple attributes incorporated in a health status measure, to the response format's ability to capture different degrees of subjectivity, and to the approach taken to summarize within and across health attributes.

Research Implications

Further exploratory work could proceed along clinical and conceptual lines. From a clinical perspective, we studied a group of individuals who spoke English, were well-educated, and, although experiencing a chronically painful condition that interferes with functional abilities, were not morbidly ill. Different results might be obtained from clinical groups of different socioeconomic status or who are facing a life-threatening illness. Among the latter, the elicitation of HYEs may not be subject to the recency effect considered here. Study designs similar to that used here would be helpful in carrying out this further exploratory work.

From a conceptual perspective, the controversy outlined in the introduction could be addressed by comparing HYEs with the results obtained using an alternative approach originally advocated by Sox and colleagues (28) and reiterated by Johannesson et al. (13). This alternative approach involves using the time trade-off technique to determine the risk-free time in good health that is equivalent to an ill health profile, and then determining the risk-based utility for that profile by referring to a utility function for time in good health derived using the certainty equivalence technique. If these two different approaches tap into individuals' composite attitudes toward time discount, risk regarding quality of life, and risk regarding survival time, their results should agree with each other. A project exploring this conceptual issue may help to resolve some of the arguments haunting the QALYs-HYEs controversy.

REFERENCES

1. Bellamy N. Pain assessment in osteoarthritis: Experience with the WOMAC osteoarthritis index. *Semin Arthritis Rheum*. 1989:18(suppl 2);14-17.
2. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: A health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *Journal of Orthopaedic Rheumatology*. 1988;1:95-108.
3. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *Journal of Orthopaedic Rheumatology*. 1988;15:1833-1840.
4. Boyle MH, Torrance GW, Sinclair JC, et al. Economic evaluation of neonatal intensive care of very-low birth-weight infants. *N Engl J Med*. 1983;308:1330-1337.

5. Buckingham K. A note on HYE (health years equivalent). *J Health Econ*. 1993;12:301-309.
6. Culyer AJ, Wagstaff A. QALYs versus HYEs. *J Health Econ*. 1993;12:311-323.
7. Culyer AJ, Wagstaff A. QALYs versus HYEs: A reply to Gafni, Birch, and Mehrez. *J Health Econ*. 1995;14:39-45.
8. Dolan P. A note on QALYs versus HYEs: Health states versus health profiles. *Int J Technol Assess Health Care*. 2000;16:1220-1224.
9. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ*. 1995;4:289-299.
10. Fowler FJ, Cleary PD, Massagli MP, Weissman J, Epstein A. The role of reluctance to give up life in the measurement of the value of health states. *Med Decis Making*. 1995;15:195-200.
11. Gafni A, Birch S, Mehrez A. Economics, health, and health economics: HYEs versus QALYs. *J Health Econ*. 1993;12:325-339.
12. Johanesson M. Quality-adjusted life years versus healthy-years equivalents: A comment. *J Health Econ*. 1995;14:9-16.
13. Johannesson M, Pliskin JS, Weinstein MC. Are healthy-years equivalents an improvement over quality-adjusted life years? *Med Decis Making*. 1993;13:281-286.
14. Llewellyn-Thomas HA. Health state descriptions: Purposes, issues, a proposal. *Med Care*. 1996;34(suppl):DS109-DS118.
15. Llewellyn-Thomas HA, Arshinoff R, et al. In the queue for total joint replacement: Patients' perspectives on waiting times. *J Eval Clin Pract*. 1998;4:63-74.
16. Llewellyn-Thomas HA, Schwartz C. Response shift effects in patients' evaluations of health states: Sources of artifact. In: Schwartz C, Sprangers M, eds. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association Press; 2000:109-122.
17. Loomes G. The myth of the HYE: QALYs versus HYEs. *J Health Econ*. 1995;14:1-7.
18. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med*. 1989;28:299-308.
19. McNeil BJ, Weichselbaum R, Pauker SG. Fallacy of the five-year survival in lung cancer. *N Engl J Med*. 1978;299:1397-1401.
20. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making*. 1989;9:142-149.
21. Mehrez A, Gafni A. The healthy-years equivalents: How to measure them using the standard gamble approach. *Med Decis Making*. 1991;11:140-146.
22. Mehrez A, Gafni A. Healthy-years equivalents versus quality-adjusted life years. *Med Decis Making*. 1993;13:287-292.
23. Nichol G, Llewellyn-Thomas HA, Naylor CD, Thiel EC. The relationship between cardiac functional capacity and patients' symptom-specific utilities for angina: some findings and methodologic lessons. *Med Decis Making*. 1996;16:78-85.
24. Nord E. *Cost-value analysis in health care: Making sense out of QALYs*. New York: Cambridge University Press; 1999:106-108.
25. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*. 1992;66:3-8.
26. Ried W. QALYs versus HYEs: What's right and what's wrong? A review of the controversy. *J Health Econ*. 1998;17:607-625.
27. Sherbourne C, Sturm R, Wells KB. Development of utility scores for the SF-12 (abstract). *Qual Life Res*. 1997;6:721.
28. Sox HC, Blatt MA, Higgins MC, et al. *Medical decision making*. Boston: Butterworths Publishers; 1988.
29. Sutherland HJ, Llewellyn-Thomas HA, Boyd NF, et al. Attitudes toward quality of survival: the concept of "maximal endurable time." *Med Decis Making*. 1982;2:299-309.
30. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ*. 1986;5:1-30.
31. Torrance GW. Utility approach to measuring health related quality of life. *J Chronic Dis*. 1987;40:593-600.

32. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programmes. *Health Serv Res*. 1972;7:118-133.
33. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*. 1974;185:1124-1131.
34. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practice. *N Engl J Med*. 1977;296:716-721.
35. Williams A. Economics of coronary artery bypass grafting. *BMJ*. 1985;291:326-329.
36. Williams JI, Llewellyn-Thomas HA, Arshinoff R, et al. The burden of waiting for hip and knee replacements in Ontario. *J Eval Clin Pract*. 1997;3:1-10.