

Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review

Pirjo Räsänen, Eija Roine

Helsinki and Uusimaa Hospital Group

Harri Sintonen

University of Helsinki and Finnish Office for Health Technology Assessment

Virpi Semberg-Konttinen

Helsinki and Uusimaa Hospital Group

Olli-Pekka Ryyänen

University of Kuopio

Risto Roine

Helsinki and Uusimaa Hospital Group

Objectives: The objectives of this study were to identify, in a systematic literature review, published studies having used quality-adjusted life years (QALYs) based on actual measurements of patients' health-related quality of life (HRQoL) and to determine which HRQoL instruments have been used to calculate QALYs. Furthermore, the aims were to characterize studies with regard to medical specialty, intervention studied, results obtained, quality, country of origin, QALY gain observed, and interpretation of results regarding cost-effectiveness.

Methods: Systematic searches of the literature were made using the MEDLINE, Embase, CINAHL, SCI, and Cochrane Library electronic databases. Initial screening of identified articles was based on abstracts read independently by two of the authors; full-text articles were again evaluated by two authors, who made the final decision on which articles should be included.

Results: The search identified 3,882 articles; 624 were obtained for closer review. Of the reviewed full-text articles, seventy reported QALYs based on actual before–after measurements using a valid HRQoL instrument. The most frequently used instrument was EuroQol HRQoL instrument (EQ-5D, 47.5 percent). Other instruments used were Health Utilities Index (HUI, 8.8 percent), the Rosser–Kind Index (6.3 percent), Quality of Well-Being (QWB, 6.3 percent), Short Form-6D (SF-6D, 5.0 percent), and 15D (2.5 percent). The rest (23.8 percent) used a direct valuation method: Time Trade-Off

The authors thank Merja Jauhiainen, Information Officer at the Finnish Institute for Occupational Health, and Iris Pasternack, MD, Occupational Health Physician, Researcher at the National Research and Development Centre for Welfare and Health (STAKES) and the Finnish Office for Health Technology Assessment (FinOHTA). The study was financially supported by the Finnish Office for Health Technology Assessment.

(10.0 percent), Standard Gamble (5.0 percent), visual analogue scale (5.0 percent), or rating scale (3.8 percent). The most frequently studied medical specialties were orthopedics (15.5 percent), pulmonary diseases (12.7 percent), and cardiology (9.9 percent). Ninety percent of the studies came from four countries: United Kingdom, United States, Canada, the Netherlands. Approximately half of the papers were methodologically high quality randomized trials. Forty-nine percent of the studied interventions were viewed by the authors of the original studies as being cost-effective; only 13 percent of interventions were deemed not to be cost-effective.

Conclusions: Although QALYs gained are considered an important measure of effectiveness of health care, the number of studies in which QALYs are based on actual measurements of patients' HRQoL is still fairly limited.

Keywords: Cost-effectiveness, Health-related quality of life, Quality-adjusted life years

Investments into health care traditionally have been made without detailed information on the health gains produced. As resources are scarce, investments should be allocated in the most cost-effective way, but without comparative effectiveness data, decision making is often on shaky ground. Especially data allowing the comparison of the effectiveness of various interventions across different medical specialties have been scarce because most comparative studies have used disease-specific outcome measures.

Allocation decisions based on clinical results only may lead to inappropriate distribution of resources regarding societal welfare. Thus, when considering various alternatives, one should, in addition to the expected gains, also take into account the lost opportunities that inevitably follow an investment decision. Resource allocation, under optimal conditions, should generate maximal benefits for the society, but especially in healthcare allocation, decisions are often combined with significant uncertainty. To be able to reliably measure the cost-effectiveness of various interventions, thus, is one of the key targets in the pursuit for good-quality, cost-conscious health care.

During recent years, it has been acknowledged that, in addition to the length of life, quality of life also is of importance. This acknowledgment has resulted in attempts to develop new, generic methods for the estimation of treatment results that also take into account patient preferences. To solve the problem of comparability of measurements, health economists have introduced the concept of health-related quality of life (HRQoL) as an indicator of individual well-being and as a potential yardstick for the estimation of health gains produced by treatments. HRQoL can be used to describe the effects of an illness on the quality of life and the effect of clinical interventions on health and general well-being (1). In addition to the disease and its treatment, HRQoL is affected by the general condition of the individual in question, other health problems and sickness experiences the person may have, the patient's phase of life, as well as the tasks and goals the patient has.

Two kinds of HRQoL instruments exist, generic and disease-specific ones. Disease-specific instruments are used for studying the most important effects of a given disease.

They are not suited, thus, for comparison of treatment results across various diseases. Their main purpose is to assist clinical decision making, and they are usually sensitive in measuring results of specific treatments. Good examples of disease-specific instruments are for instance the Knee Society Score (KSS) evaluating pain and mobility in patients with knee problems, and the Harris Hip Score (HHS) designed for the assessment of symptoms of hip disorders (11;12).

The generic instruments can be used for diverse patient groups independent of the underlying disease or disability. Generic instruments can be methodologically classified into profile and single index score measures. The former measures describe the health state from the standpoint of various physical and emotional dimensions, such as vitality, role emotional, bodily pain, general health, social function, and so on, as in the widely used Short Form-36 (SF-36) instrument. The latter produce a single index score on a 0–1 scale (although some instruments produce also negative scores), which is a necessary requirement for the calculation of quality-adjusted life-years (QALYs) used for commensurate appraisal of the cost-effectiveness of various healthcare interventions. When choosing a HRQoL instrument, special attention needs to be paid to its empirical, theoretical, and technical characteristics, such as validity, reliability, sensitivity, usability, and interpretability (1;4;7). Generic, single index score instruments include for instance the EQ-5D (EuroQol), the SF-6D (derived from RAND-36/SF-36), the HUI (Health Utilities Index Mark II/Mark III), the AQoL (Assessment of Quality of Life), and the 15D (1;13).

The QALY makes it possible to express the effectiveness of health care as a combination of a change both in the length and/or quality of life. During recent years, the QALY has been recognized as the currently most important indicator of effectiveness of healthcare interventions. This recognition is reflected, for instance, in the standpoint of the National Institute of Health and Clinical Excellence (NICE), providing national guidance on treatments and care for those using the National Health Service (NHS) in England and Wales, that it uses the QALY as its principal measure of health outcome (15). The increasing utilization of QALYs gained as a measure of effectiveness is also evidenced by the finding

that the number of references found with the search term QALY in the MEDLINE database during this decade has increased by approximately 10 percent every year.

Many of the articles reporting QALYs as end points, however, are based on economic modelling using HRQoL data obtained from many different sources or derived from healthcare professionals' estimates of the HRQoL associated with certain disease states. Such estimates, however, are likely to be biased as they represent the care providers' views, not those of patients. Consequently, it is of importance that QALY calculations are based on actual measurements of patients' HRQoL by either multiattribute (like the available generic HRQoL instruments) and/or direct (the patients' assessment and valuation of their own health status) measures. The aim of this systematic literature review was to identify articles having used patient-derived HRQoL as the basis for the QALY calculations and to characterize the studies with regard to medical specialty, intervention studied, results obtained, HRQoL instrument used, quality, country of origin, QALY gain observed, and interpretation of results regarding cost-effectiveness.

METHODS

Literature Search

Computerized literature searches were performed, without any language restrictions, using the MEDLINE (1966–June 2004), Embase (1966–June 2004), CINAHL (1982–June 2004), and Science Citation Index (1982–June 2004) databases and the Cochrane library (Issue 2, 2004). The detailed search strategy is available in the full report www.stakes.fi/finohta/e/reports/. In addition, some articles were identified by scanning reference lists of included articles, running a MEDLINE search using the name of the principal author of each included article as the search term, and consulting experts in the field of economic evaluation. Finally, we also compared the results of our search with the listing of cost-effectiveness ratios published in the Cost Effectiveness Analysis Registry (<http://www.tufts-nemc.org/cearegistry/index.html>).

Selection of Publications

Initial screening of the identified articles was based on their abstracts. All abstracts were read independently by at least two of the authors. Selection of relevant articles was based on the information obtained from the abstracts and was agreed upon in discussion between the authors. When an abstract did not give sufficiently precise information about the study or such information was not available at all, the full-text article was obtained for further review.

Full-text articles obtained for closer inspection were read independently by two of the authors (P.R., E.R., or R.R.) and placed in one of five categories according to predefined criteria (Table 1). If the two readers disagreed about the category the article belonged to, the article was read by a third person, and all three evaluators then discussed the article together to reach consensus using the criteria discussed in Table 1 and below. Included were articles that, in a scientifically valid manner, compared HRQoL of patients in a before–after setting and in which HRQoL had been assessed by the patients with a generic HRQoL instrument recognized to produce a valid single index score for the calculation of QALYs (15D, EQ-5D, SF-6D, HUI, AQoL, QWB, Rosser-Kind) or in which HRQoL had been assessed by a direct valuation method (Time-Trade-Off, Standard Gamble, visual analog scale, or rating scale).

Quality of Included Studies

The strength of evidence given in selected papers was considered with regard to the study design used and study performance as described earlier (9;10). For *study design*, scores were assigned to five classifications. Large randomized controlled trials (RCTs), defined as those with at least fifty subjects in each arm, were given a score of 5. Small RCTs had a score of 3, prospective nonrandomized studies 2, retrospective comparative studies 1, and noncontrolled series 0.

For *study performance*, five areas of interest were considered, as shown in Table 2. When reviewing a study, each of these five areas was given a score of 0, 1, or 2. A score of 0 applied when relevant information was missing or given in only minimal detail; 1 indicated that reasonable detail was

Table 1. Criteria for classification of reviewed full-text articles into one of five categories

A	HRQoL measured with a generic instrument allowing calculation of QALYs (15D, EQ-5D, SF-6D, HUI, AQoL, GWB, or Rosser-Kind), or assessed by a direct valuation method (TTO, SG, VAS, or RS) either in the study reported, or in a clearly identifiable other study, and HRQoL was assessed both before and after the intervention
B	HRQoL measured with a generic instrument allowing calculation of QALYs (15D, EQ-5D, SF-6D, HUI, AQoL, GWB, or Rosser-Kind), or assessed by a direct valuation method (TTO, SG, VAS, or RS), but HRQoL was assessed only before or after the intervention
C	HRQoL data obtained from poorly defined sources or determined with an instrument not suitable for calculation of QALYs, although the article reports QALYs using diverse transformation processes
D	HRQoL estimated (mainly or entirely) by expert panels, or volunteers, or based on literature
E	Review article not reporting original assessment of HRQoL or an inadequate or poorly defined research setting

AQoL, Assessment of Quality of Life; 15D, 15D HRQoL instrument; EQ-5D, EuroQoL HRQoL instrument; HRQoL, health-related quality of life; HUI, Health Utilities Index Mark II/Mark III; QWB, Quality of Well-Being Scale; Rosser-Kind, Rosser-Kind HRQoL instrument; RS, Rating Scale; SF-6D, SF-6D HRQoL instrument; SG, Standard Gamble; TTO, Time Trade-off; VAS, Visual analogue scale.

Table 2. Classification of study performance according to Hailey et al., 2004 (10)

Areas of interest		Points considered
1	Patient selection	Methods of randomization/selection. Equivalence of intervention and control groups. Dropouts prior to commencement of intervention
2	Description/specification of the interventions	Adequate description for both intervention and control groups
3	Specification and analysis of study	Sample size; statistical methods used; clear specification of outcome measures
4	Patient disposal	Length of follow-up; dropouts; compliance failures
5	Outcomes reported	Fullness and clarity of reporting. Missing results; statistical summary. Whether conclusions were consistent with data

provided but there were some important limitations; and a score of 2 was allocated when information provided was satisfactory, with no significant limitations. Each study, therefore, had a possible maximum score of 10 for performance.

RESULTS

Retrieved Articles

The literature search identified 4,878 publications. However, 996 were either reviews, letters, or editorials and, as we were looking for original studies, not included for further review. Furthermore, the 996 excluded articles also included publications dealing with prevention or screening, topics which had been decided to be excluded from the review. Thus, we were left with 3,882 articles potentially reporting QALYs as outcome measures. After screening of abstracts, 624 full-text articles were selected for closer inspection. Of them, seventy-two (representing seventy separate studies) were deemed to fulfill the selection criteria and were included in the review. In eighty cases (13 percent of the 624 full-text articles), the initial evaluation of the two independent reviewers differed regarding whether the article was based on clearly identifiable HRQoL data obtained with a valid instrument (groups A or B in Table 1) or not. In those cases, the article was also evaluated by a third person and the final decision was made in a consensus meeting. Of those eighty articles, eighteen were finally deemed to merit inclusion in the review. Comparison of our search result with the Cost Effectiveness Analysis Registry database identified fifty-nine additional candidate articles; none of which, however, when studied in more detail, turned out to fill the inclusion criteria of the review.

Study Classification

The seventy selected publications were grouped by the HRQoL instrument used in the study (Table 3, some studies used several instruments), by the nineteen medical specialties they represented (Table 4), and by the country of origin (Table 5). Of the included articles, 71 percent had been published in specialty journals, 20 percent in general medical journals, and 8 percent in journals mainly devoted to health economics, assessment of healthcare technologies, or health-care administration. One included study had been published

Table 3. HRQoL instrument used in the study^a

Instrument	N	% of all	%
HRQoL instruments			76
15D	2	2.5	
EQ-5D (EuroQol)	37	46.8	
HUI (Health Utilities Index – Mark II/ Mark III)	7	8.9	
QWB (Quality of Well-Being Scale)	5	6.3	
Rosser–Kind	5	6.3	
SF-6D	4	5.1	
Direct valuation			24
SG (Standard Gamble)	4	5.1	
TTO (Time Trade-Off)	7	10.1	
Rating Scale	2	3.8	
VAS	4	5.1	

^a In some studies several instruments were used. HRQoL, health-related quality of life; VAS, visual analog scale.

Table 4. Clinical specialties of the included studies

Specialty	N (%)
Orthopedics	11 (15.7)
Pulmonary disease	9 (12.9)
Cardiology	7 (10)
Neurology	6 (8.6)
Rheumatology	6 (8.6)
Otorhinolaryngology	5 (7.1)
Transplantation surgery	5 (7.1)
Psychiatry	4 (5.7)
Oncology	4 (5.7)
Gynecology	2 (2.9)
Intensive care	2 (2.9)
Urology	2 (2.9)
Endocrinology	2 (2.9)
Infectious disease	1 (1.4)
Nephrology	1 (1.4)
Dental surgery	1 (1.4)
Gastroenterology	1 (1.4)
General surgery	1 (1.4)

Table 5. Country of origin of the studies

Country	N (%)
United Kingdom	23 (32.9)
The Netherlands	18 (25.7)
Canada	11 (15.7)
USA	11 (15.7)
Spain	2 (2.9)
Germany	1 (1.4)
Hong Kong	1 (1.4)
Norway	1 (1.4)
Sweden	1 (1.4)
Multinational	1 (1.4)

as a dissertation. Sixty-seven of the articles were in English, one in Norwegian, one in Dutch, and one in Spanish.

A total of 31 percent of the studies mainly were concerned with pharmacological therapy and 26 percent with surgical interventions. The rest covered various types of conservative treatment, rehabilitation, diagnostic imaging or preventive services. The interventions studied covered a wide range from transplantation surgery to spa-exercise therapy. The most commonly studied interventions were treatment of coronary heart disease, total hip arthroplasty, and cochlear implant with four studies concerned with each of them. An economic analysis was present in 86 percent of the studies; nine studies reported only HRQoL and QALY results.

Further details of each study in terms of clinical specialty, intervention, aim, data used, method used, perspective of economic analysis, cost data used, results concerning HRQoL assessment, number of and cost per QALYs gained by intervention, quality of study, and any methodological or other limitations are available in the full report at www.stakes.fi/finohta/e/reports/. Two of the studies were the subject of more than one paper; in those cases the results of separate articles were combined in the table.

Study Design and Quality

Qualitative analysis showed that approximately half of the articles were based on randomized controlled trials. Also most of the remaining studies were comparative. Study performance on the scale of 0–10 (Table 2) was considered good (8–10 points on the scale) in 49 percent of the studies, fair (6–7 points) in 29 percent of the studies, and fair to poor (4–5 points) in 22 percent of the studies. None of the studies was deemed to be of poor quality (<4 points on the scale). Four studies used economic modeling: three the Markov model, and one a decision analytic model.

Reported Number of QALYs and Costs per QALY

The reported number of QALYs gained by various treatments varied widely from negative to eight, depending on the intervention studied and, partly, on over how many years the QALY gain was extrapolated. Also, the cost per QALY

showed great variation, from less than €1,000/QALY to over €1,000,000/QALY.

Study Conclusions

Apart from reporting QALY results, most of the studies also discussed the cost per QALY in terms of acceptability for society. Forty-nine percent of the studied interventions were viewed by the authors of the original studies as being cost-effective; only 13 percent of the interventions were deemed not to be cost-effective.

DISCUSSION

Although QALY is considered an important measure of effectiveness of health care, only a fairly limited number of studies are really based on actual measurements of patients' HRQoL. In many studies identified during the review process, HRQoL data were obtained from vaguely defined sources or estimated by healthcare professionals. Even though healthcare professionals certainly are aware of the clinical nature of a disease and the burden it can cause for their patients, it is unlikely that they—having never experienced the disease themselves—would really be able to judge patients' HRQoL properly. This unlikelihood is evidenced by a recent study in which correlations between prostate cancer patients' and clinicians' utilities were very low and not statistically significant (2) and also in line with some earlier findings (3;16;17). Therefore, studies based on real patient data are probably of much more value for the decision maker pondering allocation of resources.

On the other hand, we identified several studies in which HRQoL had been studied in a proper before–after setting but which did not include the term QALY in their reports. Most of those studies would have provided a possibility to calculate also QALYs, but as they did not, they were not included, as we could not be sure whether we would have covered them extensively with the search strategy specifically designed to identify studies reporting QALYs. It is possible that studies in which the effect of an intervention on HRQoL is absent or minimal are more likely not to report QALYs as those with a more positive result and that, therefore, the studies included in our review may give a somewhat biased and overoptimistic impression about the QALY gains that can be achieved by medical interventions.

Our approach differs from earlier reviews as we were explicitly looking only for studies in which the reported QALYs were based on a before–after assessment of HRQoL with a valid instrument. Consequently, it is not possible to compare—at least in a very productive manner—our results to those of earlier reviews having used different inclusion criteria. We acknowledge that it is not always possible to obtain valid HRQoL data directly from the patients (for instance in children, in demented patients, and so on), and in such instances, HRQoL must be estimated using proxies or other methods. Although such studies were excluded

from the present analysis, it does not mean that their results would not be valuable. Neither does it suggest that earlier attempts to systematically review such studies and to describe their results, such as for example, the comprehensive Cost Effectiveness Analysis Registry (<http://www.tufts-nemc.org/cearegistry/index.html>), would be of lesser importance than our approach. However, in most cases, patients can certainly independently rate their HRQoL themselves. We believe that, in such cases, the use of estimated data may be misleading and, therefore, wanted to scrutinize the existing scientific literature reporting QALYs to be able to list and describe the studies that really are based on patient-derived data.

Study quality was in the majority of cases fair to good, and approximately half of the studies were based on randomized controlled trials. The results of the studies, thus, in most cases can be considered reliable and may have a direct influence on decision making. Previous studies have indicated continuing variation in the quality of cost–utility studies (5;6). Although the main emphasis of our review was not on the quality of the included studies and our approach to quality assessment, therefore, was not as stringent as that of Gerard et al. (5;6), our results may indicate that there has been slow improvement in the quality and reporting of cost–utility analyses over the years. This improvement is in agreement with the recent results of Neumann et al., who comparing articles published in 1998 to 2001 with those published in 1976 to 1997, reported an improvement in almost all quality categories assessed (14).

Of the HRQoL instruments used in the reviewed studies, the EQ-5D was by far the most popular. Although the instrument has limitations, for instance that it can only define 243 health states, it is simple to use. Furthermore, it has been developed in cooperation with several health economist groups in different countries with a strong representation from the United Kingdom and the Netherlands, which probably has furthered its adoption into research practice in those countries, from which most of the reviewed studies indeed come from.

In studies where more than one generic HRQoL instrument had been used concurrently, the number of QALYs gained differed fairly much, depending on the instrument used. This is certainly a limitation for the use of QALY data based on different instruments for comparison of the effectiveness of interventions. Another limitation preventing meaningful comparisons regarding the utility of various interventions is that there is great variation in the way the QALY results are expressed especially regarding the extrapolation of the results over time. In some articles, the time horizon on which the QALY calculation was based on was less than a year; in some others several years or until death. At the same time, the comparison of cost–utility results is hampered by variation in the costing methods and the many different perspectives (e.g., societal or healthcare provider perspective) from which cost–utility is considered. Conse-

quently, to make the results of cost–utility analyses more valuable, there clearly is a need for common, widely agreed methodology.

Almost half of the studied interventions were considered acceptable for society in terms of cost per QALY. The threshold used for society's willingness to pay per QALY, however, varied, indicating that, at the moment, there is no universally accepted standard for an appropriate threshold for resource-allocation decisions. In this respect, our results are in line with a recent report about the prevailing judgments about society's willingness to pay for a QALY (8).

CONCLUSIONS

Although QALYs gained are considered an important measure of effectiveness of health care, the number of studies reporting QALYs based on actual measurement of patients' HRQoL is still fairly limited. Such studies, however, are urgently needed to ensure that allocation of healthcare resources is based on scientific evidence on the value of various interventions regarding their ability to produce societal welfare.

CONTACT INFORMATION

Pirjo Räsänen, MSc, RN (pirjo.rasanen@stakes.fi), Researcher, **Eija Roine** (eija.roine.0392@student.uu.se), Group Administration, Helsinki and Uusimaa Hospital Group, P.O. Box 100, 00029 HUS, Finland

Harri Sintonen, PhD (harri.sintonen@helsinki.fi), Professor of Health Economics, Department of Public Health, University of Helsinki, P.O. Box 41, 00014 Helsinki, Finland; Finnish Office for Health Technology Assessment, National Research and Development Centre for Welfare and Health, P.O. Box 220, 00531 Helsinki, Finland

Virpi Semberg-Konttinen, MSc, RN, MJD (virpi.semberg-konttinen@hus.fi), Officer for Audit Board, Group Administration, Helsinki and Uusimaa Hospital Group, P.O. Box 100, 00029 HUS, Finland

Olli-Pekka Ryyänen, MD, PhD (ollipekka.ryynanen@uku.fi), Professor, Department of Public Health and General Practice, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland

Risto Roine, MD, PhD (risto.p.roine@hus.fi), Chief Physician, Group Administration, Helsinki and Uusimaa Hospital Group, P.O. Box 100, 00029 HUS, Finland

REFERENCES

1. Drummond MF, O'Brien BJ, Stoddard GL, Torrance GW. 1997. *Methods for the Economic evaluation of health care programmes*. Oxford: Oxford University Press.
2. Elstein AS, Chapman GB, Chmiel JS, et al. Agreement between prostate cancer patients and their clinicians about utilities and attribute importance. *Health Expect*. 2004;7:115-125.

3. Epstein AM, Hall JA, Tognetti J, Son LH, Conant L. Using proxies to evaluate quality of life. *Med Care*. 1989;27(Suppl):91-98.
4. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2:i-iv,1-74.
5. Gerard K. Cost-utility in practice: A policy maker's guide to the state of the art. *Health Policy*. 1992;21:249-279.
6. Gerard K, Smoker I, Seymour J. Raising the quality of cost-utility analyses: Lessons learnt and still to learn. *Health Policy*. 1999;46:217-238.
7. Gold MR, Russell LB, Siegel JE, Weinstein MC. 1996. *Cost-effectiveness in health and medicine*. Oxford: Oxford University Press.
8. Greenberg D, Winkelmayr WC, Neumann PJ. Prevailing judgments about society's willingness to pay for QALY or life-year gained. *Ital J Public Health*. 2005;2(Suppl 1):301.
9. Hailey D, Ohinmaa A, Roine R. Evidence for the benefits of telecardiology applications: A systematic review. *Alberta Heritage Found Med Res*. 2004.
10. Hailey D, Ohinmaa A, Roine R. Study quality and evidence of benefit in recent assessments of telemedicine. *J Telemed Telecare*. 2004;10:318-324.
11. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am*. 1969;51:737-755.
12. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res*. 1989;248:13-14.
13. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol*. 2003;56:317-325.
14. Neumann PJ, Greenberg D, Olchanski NV, Stone PW, Rosen AB. Growth and quality of the cost-utility literature, 1976-2001. *Value Health*. 2005;8:3-9.
15. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgements. *BMJ*. 2004;329:224-227.
16. Sintonen H. *The 15D-measure of health-related quality of life. I. Reliability, validity and sensitivity of its health state descriptive system*. National Centre for Health Program Evaluation, Working Paper 41, Melbourne 1994. Available at: <http://www.buseco.monash.edu.au/centres/che/publications.php#rp>.
17. Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM. Who should measure quality of life, the doctor or the patient? *Br J Cancer*. 1989;57:109-112.