QUALITY-ADJUSTED LIFE-YEARS FOR THE ESTIMATION OF EFFECTIVENESS OF SCREENING PROGRAMS: A SYSTEMATIC LITERATURE REVIEW

Suvi Mäklin

National Institute for Health and Welfare, Finnish Office for Health Technology Assessment
email: suvi.maklin@thl.fi

Pirjo Räsänen

National Institute for Health and Welfare, Finnish Office for Health Technology Assessment;
Hospital District of Helsinki and Uusimaa

Riikka Laitinen, Niina Kovanen

National Institute for Health and Welfare, Finnish Office for Health Technology Assessment

Ilona Autti-Rämö

The Social Insurance Institution

Harri Sintonen

University of Helsinki

Risto P. Roine

Hospital District of Helsinki and Uusimaa

Objectives: The aim of this study was to identify and characterize studies that have used quality-adjusted life-years (QALYs) based on measurements of patients' health-related quality of life (HRQoL) as an indicator of effectiveness of screening programs.

Methods: Systematic search of the literature until March 2010, using several electronic databases. Initial screening of articles based on abstracts, and evaluation of full-text articles were done by at least two of the authors.

Results: The search identified 1,610 articles. Based on review of abstracts, 431 full-text articles were obtained for closer inspection and, of these, 81 reported QALYs based on patient-derived data using a valid HRQoL assessment. The most frequently used method to assess HRQoL was Time Trade-Off (55 percent) followed by EQ-5D (26 percent). The most frequently studied medical conditions were malignant diseases (23 percent) followed by cardiovascular diseases (19 percent). All studies employed some kind of modeling with the Markov model being the most prevalent type (65 percent). Majority of the articles (59 percent) concluded that the screening program studied was cost-effective.

Conclusions: The use of QALYs in the evaluation of screening programs has expanded during the last few years. However, only a minority of studies have used HRQoL data derived from patients, using direct or indirect valuation. Further investigation and harmonization of the methodology in evaluation of screening programs is needed to ensure better comparability across different screening programs.

Keywords: Screening, HRQoL (health-related quality of life), QALY (quality-adjusted life-years), Systematic review

Screening aims to identify people at high risk for a disease or those who already have the disease in question. A prerequisite for a successful screening program is that early diagnosis and treatment can improve the natural prognosis of the disease, and that appropriate treatment can be guaranteed to all people diagnosed with the disease. Screening programs profoundly differ from the situation where the patient seeks care due to symptoms. The WHO set ten criteria for a good screening program in the 1960s (19) and those criteria have been further developed into more comprehensive ones used today (e.g., 12;18). The increased health-related knowledge of people raises demand for quality control and this applies also to screening programs that need to be shown to be effective (1).

The potential of health care to influence the natural course of different diseases has increased during the last decades. However, at the same time the costs of health care have multiplied and threaten to exceed available resources. Thus it is important to direct healthcare resources so that best possible gain in health is achieved, and that the gain can be produced at reasonable costs (7). Health technology assessment and economic evaluation, therefore, are of paramount importance when trying to find interventions that are of high quality and cost-effective. This is also true in the case of screening.

It is acknowledged that the length of life is not the only reasonable outcome when assessing success of healthcare interventions, but also its quality is of importance. Thus instruments to measure patients' own perceptions of their well-being have been developed, and, to assure the comparability of results, the concept of health-related quality of life (HRQoL) has been introduced. HRQoL describes the effect of a disease to patient's quality of life and the effects of a clinical intervention on his or her health and general well-being. Quality of life is nevertheless influenced not only by the disease and its treatment, but also the individual's living conditions, other possible health problems, own experience of the disease, life situation, and tasks

and goals (7). Both, generic and disease-specific quality of life can be measured. Disease-specific instruments generally do not allow a meaningful comparison between different diseases. By contrast, generic measures can be used to compare the effectiveness across various patient groups and medical specialties.

The use of HRQoL data in economic evaluation requires a measure that not only combines quality and quantity of life, but also incorporates a value on health states. These utility weights, can then be used in calculation of quality-adjusted life-years (QALY) and for comparing the cost-effectiveness of healthcare interventions (7;11). The utility weights can be obtained either by direct valuation of health states (using the Standard Gamble, Time-Trade-Off, Rating Scale, Visual Analogue Scale) or by using one of the pre-scored multi-attribute health status classification systems (e.g., Quality of Well-Being, EuroQol-5D, 15D). Methodologically the generic HRQoL instruments are classified into those producing a profile- and/or one index number. Profile measures, for example the widely used SF-36, describe health states through different physical and mental parameters. The single index HRQoL instruments combine the answers to individual questions into a single index number (usually ranging between 0 and 1). For preference-based measures, the single index number is taken from a scoring algorithm, which is based on a pre-existing set of utility weights by the general population. QALYs, are a recognized way to measure the effectiveness of healthcare interventions and are used widely in various countries. In the UK for instance, the National Institute for Health and Clinical Excellence (NICE), uses QALYs as its main measure of effectiveness (13;15).

The aim of this systematic literature review was to examine and describe the use of QALYs based on patient-derived preference-based QoL data in evaluation of screening programs. Patient-derived HRQoL data may be obtained by direct valuation or by indirect valuation (patients filling in one of the pre-scored generic HRQoL questionnaires). This article also aims to characterize the identified screening studies with regard to clinical specialty, country of origin, aim of the study, target group, subjects, methods, HRQoL instrument employed, perspective, costs, QALY or cost/QALY gain observed, authors' conclusions, quality of the study, and limitations (Supplementary Table 2, which can be viewed online at www.journals.cambridge.org/thc2012016). The approach is similar to our earlier publication about the use of QALYs in evaluation of various medical interventions (16).

METHODS

Literature Search

Computerized literature searches were performed, without any language restrictions, using the Medline (1966- March 2010), Embase (1966- March 2010), CINAHL (1982- March 2010), and Science Citation Index (1982- March 2010) databases and the Cochrane library (Issue 2, 2004; Issue 4, 2007; Is-

sue 1, 2010). The detailed search strategy is available on the Supplementary Table 4, which can be viewed online at www.journals.cambridge.org/thc2012018.

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 sufficient information about the study, the full-text article was obtained for further review. Full-text articles obtained for closer inspection were read independently by two of the authors (S.M., P.R., R.S., N.K., or R.R.). 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.

Included were original articles that used QALYs in assessing effectiveness of screening. Screening programs had to be directed at diseases that were clearly identified in the article. Included were both, universal screening programs, as well as screening targeted at groups of individuals who had a higher than average risk to be taken ill with the screened disease. Furthermore, the calculation of QALYs had to be based on utilities estimated by patients using a valid instrument; either a generic HRQoL measure (15D, EQ-5D, SF-6D, HUI, AQoL, QWB, Rosser-Kind), or a direct valuation method (TTO, SG, VAS, or RS). In the case of antenatal and neonatal screening programs, the utilities were based on parents' utilities, or on the utilities of children or teenagers living with the condition screened. Articles using utilities based on population data or expert opinion were excluded.

Quality of Included Studies

As great majority of the included studies were economic evaluations, the methodological quality of the papers was assessed using a checklist for assessing economic evaluations (7). Quality assessment was done only as a descriptive measure; no studies were excluded due to low quality.

RESULTS

Retrieved Articles

The literature search identified 1,610 articles (Figure 1). Most of the articles were in English, but also articles in Dutch, German, and Japanese were examined. Of the identified articles, 147 were reviews, letters, or editorials and, as we were looking for original studies, not included for further review. After removing duplicates, 846 articles remained, and based on review of their abstracts, 431 were ordered for full-text evaluation. Of them eighty-one fulfilled the selection criteria and were included in the review. In twenty-five cases (6 percent of the 431 full-text articles), the initial evaluation of the two independent reviewers
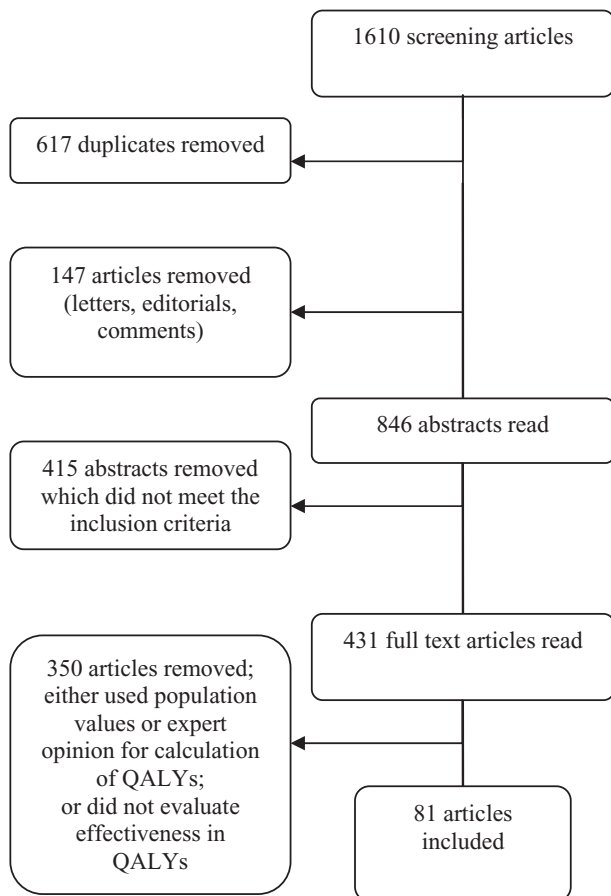
Figure 1. Number of studies identified at each step.

Table 1. Number of HRQoL Instrument Used in the Study (in Some Studies Several Instruments Were Used)

Instrument	N
HRQoL instruments	
EQ-5D (EuroQol)	21
QWB (Quality-of-Well-Being Scale)	12
HUI (Health Utilities Index)	11
Rosser-Kind	2
15D	1
AQoL	1
Direct valuation	
TTO (Time Trade-Off)	44
SG (Standard Gamble)	19
Rating Scale	2
VAS	4

differed regarding whether the article filled the inclusion criteria or not. In those cases the article was also evaluated by a third person and the final decision was based on a consensus. Most of the disagreements were due to difficulties in finding out from whom and by how the utilities were obtained. Two publications (4;10) reported results from the same analysis. Both of these can be found in the Supplementary Tables 1 (which can be viewed online at www.journals.cambridge.org/thc2012015) and 2, but are counted only once in the tables presenting study characteristics.

Study Classification

The eighty-one selected publications were grouped by the HRQoL instrument employed in the study (Table 1). The use of EQ-5D -instrument has been increasing in the recent years; 62 percent of the studies that report using EQ-5D as HRQoL instrument have been published during the past 3 years (2007–2010). Approximately 51 percent of the publications used more than one HRQoL instrument; 39 percent used two instruments, and 13 percent used three or more instruments. 16 studies included both direct and indirect methods for estimating HRQoL; that is, utilities for different health states in one model were based on different methods and/or instruments.

The publications were also sorted by the nine clinical specialties they represented (Table 2). The most often covered conditions were malignant diseases (24 percent of all included studies) and cardiovascular diseases (19 percent). Fourteen percent of the studies were concerned with contagious diseases and 13 percent with antenatal and childhood screening. The conditions screened in the included studies are listed in Table 2 according to different clinical specialties. The screening programs were targeted on the general population in forty-two articles (52 percent), populations at increased risk in seventeen (21 percent) and patients with a disease in twenty-two (27 percent), respectively.

Thirty-seven of the articles came from the United States (46 percent), twenty-one from UK (26 percent), five both from Japan and Canada (6 percent), four from the Netherlands (5 percent), two from Australia, and two from France, and one from Belgium, Finland, Italy, Sweden, and Taiwan each (1 percent).

Altogether 64 percent of the included articles had been published in specialty journals, 15 percent in general medical journals, and 21 percent in journals mainly devoted to health economics, assessment of healthcare technologies, or healthcare administration. All included articles were in English.

As effects of screening appear typically after a rather long time horizon, the evaluation of screening programs requires usually some form of economic modeling, which was also seen in this review (Table 3). Two studies compared costs and outcomes without any formal model, and in one study there was no clear description of the model used. Most of the articles (68 percent) analyzed costs and effects over the expected life time of the population to be screened, and other defined time horizons were typically long as well (10–50 years). Only one study used a short, 1 year time horizon in the base case analysis, but performed analysis also using a time horizon of 6 and 11 years (2).

Table 2. Clinical Specialities of the Included Studies and Diseases Within the Clinical Specialities.

Speciality	N (%)	Diseases or methods
Antenatal and childhood	10 (13)	Amblyopia, cystic fibrosis, group B streptococcal disease, Down syndrome, congenital adrenal hyperplasia, congenital hypothyroidism, and disorders of inborn errors of metabolism like phenylketonuria, biotinidase deficiency, maple syrup urine disease, galactosemia, homocystinuria, medium chain acyl-CoA dehydrogenase deficiency
Cardiovascular	15 (19)	Asymptomatic carotid stenosis, stroke, candidates for extracranial or intracranial bypass, cardiovascular events, carriers of factor V Leiden and prothrombin G20210A, thromboembolic complications associated with femoral catheters, depressed left ventricular function, CAD in asymptomatic diabetic patients, atrial fibrillation, abdominal aortic aneurysm, alphaadducinGly460Trp variant among hypertensive patients, deep vein thrombosis, asymptomatic intracranial aneurysms
Contagious diseases	11 (14)	HIV, hepatitis C and B, *Chlamydia trachomatis*, West Nile virus
Endocrinology	9 (11)	Osteoporosis, hip fracture, bone densitometry, overweight, vertebral deformity (osteoporosis), impaired glucose tolerance, type 2 diabetes, chronic kidney disease, nephropathy
Malignant diseases	19 (24)	Oral, esophageal, breast, lung, cervical cancer, melanoma, and hepatocellular carcinoma
Nephrology	4 (5)	Prevention of polyoma virus nephropathy, condition of donor kidney transplant, nephropathy
Ophthalmology	9 (11)	Eye diseases in patients with type 2 DM, diabetic retinopathy, age-related macular degeneration, strabismus, open-angle glaucoma
Otorhinolaryngology	2 (3)	Hearing loss
Substance abuse	1 (1)	Unhealthy alcohol use

Table 3. Type of Study

Type of model	N (%)
Decision tree	12 (15)
Markov model	52 (65)
A decision tree combined with Markov model	9 (11)
Individual sampling model	1 (1)
Monte Carlo simulation	3 (4)
No structurized model	2 (3)
No clear description of model	1 (1)

Together with long time horizon, discounting of future costs and benefits was clearly stated in the vast majority of the studies (94 percent). Most of the articles used the same discount rate for both, costs and benefits, and the discount rate of 3 percent was the most frequently used. If different discount rates were used for costs and benefits, costs were discounted with a higher rate than benefits (3.5 percent or 6 percent versus 1.5 percent) (e.g., 5;17).

Most of the included articles used HRQoL data for QALY calculation from previously published studies. In many cases, a reference of a reference needed to be obtained to find out whether HRQoL data were based on values from patients', population, or expert opinion.

Only three studies focused solely on QALYs and included no cost data. The rest were cost-utility analyses, performed from the societal perspective in 48 percent, healthcare system perspective in 38 percent and third-party payer perspective in 10 percent. In one article, the chosen perspective was that of a health management organization (HMO).

The publication of evaluations of screening programs using QALYs as the measure for effectiveness has expanded in recent years. Thirty-eight percent of the references identified in the literature search were published between years 2007–2010. During these years the number of references was 617, while the corresponding figure for the previous years (1966–2006) was 993. Of the included articles, the earliest was published in 1997 in the United States.

Details of each of the included studies are available in the Supplementary Tables 1 and 2. Abbreviations used in Supplementary Table 2 are explained in Supplementary Table 3, which can be viewed online at www.journals.cambridge.org/thc2012017.

Quality of the Studies

The quality of the eighty-one economic evaluations included in this overview is summarized in the Supplementary Table 2. The studies were mainly of very high quality; most of the studies satisfied nine of ten criteria.

Reported Outcomes

The studies included in the review reported a diversity of outcomes, which makes the results difficult to compare. Some reported only QALYs gained while others reported cost per QALY or incremental cost per QALY. Hence drawing conclusions about the overall cost-effectiveness of screening programs is not relevant, but the results reported in each study are shown in the Supplementary Table 2. The screening program in question was concluded to be cost-effective in 48 (59 percent) of the included studies, when examining the conclusions drawn by

the authors in each of the original articles. Furthermore, in 20 articles (25 percent) the conclusions were more cautious, but yet reported cost-effectiveness under certain assumptions. Only ten studies (12 percent) concluded that the screening program studied was not cost-effective.

DISCUSSION

The evaluation of effectiveness and cost-effectiveness of screening is essential when deciding whether to start a new screening program, or to expand an ongoing one. The QALY provides a logical tool for the comparison of the effectiveness of different healthcare interventions and, therefore, a systematic insight into the use of QALYs in the evaluation of the effectiveness of screening programs was, in our opinion, needed. Furthermore, our present results provide a possibility to compare the published evaluations of screening programs with other healthcare interventions reviewed in our earlier article (16).

Compared with our previous review, some differences can be seen between use of QALYs in evaluation of screening and other healthcare interventions. For screening studies, malignant diseases were the most common clinical specialty, and no studies were included from the orthopedics or pulmonary disease categories, which were the most common specialties in our previous review. Furthermore, antenatal and childhood, ophthalmology and contagious diseases, were highly represented in this screening review and more or less missing from the previous one. Three topics seemed to be of most interest and each of them was included in five studies; BRCA1/2 mutation, hepatitis C and osteoporosis. The settings, target populations and research questions on these studies varied, and hence no overall conclusions about the cost-effectiveness can be drawn.

Our findings confirm that screening studies focusing solely on HRQoL or QALYs, without inclusion of cost data and cost-utility analysis, are rare. Cost-utility analyses based on calculation of QALYs, however, have been widely used for the evaluation of screening programs, though the number of studies reporting the cost-utility of screening using QALYs based on measurement of patients' HRQoL is still fairly limited.

One of our inclusion criteria was that the utility weights were elicited from patients, either by direct valuation or indirectly, using a generic HRQoL instrument. This criterion was based on the argument that only patients can give a realistic preference over the health states related to the disease (14). Some do however argue that because the general public bears the costs and also experience the consequences of healthcare decisions, their preferences should be used in the decision-making instead of patients' (see e.g., Gold et al.) (9). It is acknowledged that patients tend to give higher HRQoL values than general public and, that patients' preferences may be affected by adaptation to ill health. But in reality it is difficult to draw a line between patients and public because most people have experienced some form of ill health and are thus able to imagine the effects of certain

illness to their quality of life (6). In evaluation of screening also the effects of the screening itself on the quality of life should be better examined. For example, abnormal and false-positive screening results have been shown to have a negative impact on many psychosocial dimensions (3).

Patient-derived HRQoL data as inclusion criteria proved to be rather demanding, as the data sources in many studies were not clearly described. Previously published values were often used and it was not always clearly mentioned, whether they were based on population or patients' preferences or on expert opinion. Many articles described typical modeling studies incorporating data from various sources. Thus several studies used utilities based on both direct and indirect valuation methods, or on different generic instruments for different health states in a single model. The comparability of different instruments is not good, and the interpretation of results from such studies may be complex. The methodological differences and non-transparent reporting weaken the comparability of QALY and cost-utility results. The European Network for Health Technology Assessment (EUnetHTA Collaboration) has taken a major task in trying to unify the methodology in conducting HTA assessments including evaluation of cost-effectiveness of screening programs (8). Common methodology and respect for patient perspective will improve transparency and transferability of the results. These are critical aspects, as generalizability of economic evaluations is always limited—due to differences in healthcare systems, current practices, prevalence of the diseases, costs, and population values—and clear description of the methods and data inputs used assists in evaluating whether the results apply into one's own setting.

There are some unique characteristics and methodological issues, which have to be taken into account when evaluating screening programs with regards to, for example, the time horizon and the existence of lifelong health implications. The screening programs are usually associated with significant investments and can have major impact on the organization of health care. The screening program should be evaluated as a whole, including not only the screening test but also the invitations to screening, further examinations, and possible treatments. The difficulty in assessing realistic costs and outcomes of screening programs is compounded by the lack of outcome data and clinical trial data. Therefore in most cases, modeling with parameters from various different sources of information is necessary. Using a long time horizon, discounting is needed and hence preventive healthcare interventions may not seem as cost-effective as other interventions in the healthcare sector. Surprisingly still, only ten (13 percent) of the included studies concluded that the analyzed screening was not cost-effective while in others at least moderate cost-effectiveness was reported. The conclusions in each of the included articles are of course context-specific and depend on the healthcare system in question; different countries are willing to pay

different amounts for a QALY gained, and differences also exist in whether this has been explicitly stated or not.

Results of economic evaluation of screening programs depend on the perspective, which, in the ideal situation, is the broadest possible. Half of the included studies used the societal perspective and included a great variety of relevant costs and consequences into the analysis regardless of when and on whom they fall on. The studies with a conclusion that the screening was not cost-effective did not differ from the others; these studies dealt with a range of clinical specialties, used different HRQoL instruments and were mainly of very high quality. Half of these studies evaluated a screening program targeted to general population, using a healthcare providers' perspective.

Before deciding on national implementation of screening programs, also other factors like ethical, social and psychological aspects have to been taken into account. Evaluation of screening programs requires multi-professional teamwork, and economic evaluation—often using modeling—constitutes an essential part of this work. This article provides a review of methods and approaches used in the literature on evaluation of screening programs, with the aim to provide information on how studies on screening have estimated health-related quality of life and calculated QALYs. This information is crucial when deciding on which outcome instruments to use in the assessment of screening programs. The methods of measuring HRQoL and economic evaluation are well established and can also be adopted into evaluation of preventive healthcare interventions. While most of the studies included in the review were of high quality, there were a lot of variation in which outcome measures were chosen and incremental analysis was not included in all. The methods of economic evaluation, with regards to the unique characteristics of screening interventions, need to be examined and further developed.

CONCLUSIONS

The use of QALYs in the evaluation of screening programs has expanded during the last few years. However, only a minority of studies have used HRQoL data derived from patients. Further investigation and harmonization of the methodology in evaluation of screening programs is needed to ensure better comparability across different screening programs.

SUPPLEMENTARY MATERIAL

Supplementary Table 1
www.journals.cambridge.org/thc2012015
Supplementary Table 2
www.journals.cambridge.org/thc2012016
Supplementary Table 3
www.journals.cambridge.org/thc2012017
Supplementary Table 4
www.journals.cambridge.org/thc2012018

CONTACT INFORMATION

Suvi Mäklin, MSc, Research Officer, National Institute for Health and Welfare, Finnish Office for Health Technology Assessment, Helsinki Finland

Pirjo Räsänen, PhD, RN, Senior Researcher, Docent, National Institute for Health and Welfare, Finnish Office for Health Technology Assessment, and Hospital District of Helsinki; and Hospital District of Helsinki and Uusimaa, Group Administration, Research and Development, Helsinki, Finland

Riikka Laitinen, MSc, Development Manager, Niina Kovanen, MSc, Development Manager, National Institute for Health and Welfare, Finnish Office for Health Technology Assessment, Helsinki Finland

Ilona Autti-Rämö, MD, Chief of Health Research, Research Professor, The Social Insurance Institution, Helsinki, Finland

Harri Sintonen, PhD, Professor of Health Economics (emeritus), The Hjelt Institute, Department of Public Health, University of Helsinki, Finland

Risto P. Roine, MD, PhD, Chief Physician, Hospital District of Helsinki and Uusimaa, Group Administration, Research and Development, Helsinki, Finland

CONFLICTS OF INTERTEST

Harri Sintonen has received funding as a Board member or consultant from several drug companies (MSD, Merck, Eli-Lilly, Pfizer, and Novartis). He is the developer of the 15D and one of the developers of the EQ-5D, and receives royalties from the electronic version of the 15D. The other authors report they have no potential conflicts of interest.

REFERENCES

1. Autti-Rämö I, Mäkelä M, Sintonen H, et al. Expanding screening for rare metabolic disease in the newborn: An analysis of costs, effect and ethical consequences for decision making in Finland. *Acta Paediatr*. 2005;94:1126-1136.
2. Bamford J, Fortnum H, Bristow K, et al. Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen. *Health Technol Assess*. 2007;11:1-168.
3. Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. *Health Qual Life Outcomes.* 2007;5:3.
4. Burr JM, Mowatt G, Hernández R, et al. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: A systematic review and economic evaluation. *Health Technol Assess*. 2007;11:iii-iv, 1-190.
5. Castelnuovo E, Thompson-Coon J, Pitt M, et al. The cost-effectiveness of testing for hepatitis C in former injecting drug users. *Health Technol Assess*. 2006;10:1-93.
6. Dolan P. Whose preferences count? *Med Decis Making.* 1999;19:482-486.
7. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL, *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford University Press; 2005.
8. European Network for Health Technology Assessment. *The mission of the EUnetHTA Collaboration*. http://www.eunethta.net/Public/About_EUnetHTA/EUnetHTA-Mission/www.eunethta.net.

9. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine*.Oxford: Oxford University Press; 1996.

10. Hernandez RA, Burr JM, Vale LD. Economic evaluation of screening for open-angle glaucoma. *Int J Technol Assess Health Care*. 2008;24:203-211.

11. 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.

12. Ministry of Social Affairs and Health. *Screening programmes. A handbook for municipal authorities. Helsinki 2007*. Finland: Publications of the Ministry of Social Affairs and Health; 2007:5.

13. National Institute for Health and Clinical Excellence (NICE). *Guide to the methods of technology appraisal*. London: NICE; 2008.

14. Nord E, Pinto JL, Richardson J, Menzel P, Ubel P. Incorporating societal concerns for fairness in numerical valuations of health programs. *Health Econ.* 1999;8:25-39.

15. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgements. *BMJ.* 2004;329:224-227.

16. Räsänen P, Roine E, Sintonen H, et al. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. *Int J Technol Assess Health Care*. 2006;22:235-241.

17. Stein K, Dalziel K, Walker A, et al. Screening for Hepatitis C in injecting drug users: A cost utility analysis. *J Public Health*. 2004;26:61-71.

18. UK National Screening Committee. Programme appraisal criteria. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. http://screening.nhs.uk/criteria.

19. Wilson JMG, Jungner G. *Principles and practice of screening for disease*. Geneva: World Health Organization; 1968.