

Pirjo Räsänen

# Routine Measurement of Health-Related Quality of Life in Assessing Cost-Effectiveness in Secondary Health Care

Academic Dissertation  
University of Helsinki  
and  
Hospital District of Helsinki and Uusimaa



National Research and Development Centre  
for Welfare and Health

Research Report 163

*Editorial board*

Matti Heikkilä, chairman

Marjatta Bardy

Marko Elovainio

Mika Gissler

Riitta Haverinen

Timo Tuori

Matti Virtanen

The research presented in this series has been approved for publication after undergoing a formal referee evaluation process.

© Author and STAKES

*Language editing:* Mark Phillips

*Cover design:* Harri Heikkilä

*Layout:* Christine Strid

ISBN 978-951-33-1981-6

ISSN 1236-0732

Gummerus Printing

Vaajakoski 2006

Finland

***Supervised by***

*Professor Harri Sintonen*

*University of Helsinki*

*Docent Risto P. Roine*

*Hospital District of Helsinki and Uusimaa*

***Official opponent***

*Professor Pekka Rissanen*

*University of Tampere*

***Reviewed by***

*Professor Martti Kekomäki*

*University of Helsinki*

*Research Professor Markku Pekurinen*

*National Research and Development*

*Centre for Welfare and Health*

## Abstract

Räsänen Pirjo. Routine Measurement of Health-Related Quality of Life in Assessing Cost-effectiveness in Secondary Health Care. STAKES, Research Report 163. Helsinki, Finland, 2007. ISBN 978-951-33-1981-6

**Aims:** The aims of this study were 1) to identify and describe health economic studies that have used quality-adjusted life years (QALYs) based on actual measurements of patients' health-related quality of life (HRQoL); 2) to test the feasibility of routine collection of health-related quality of life (HRQoL) data as an indicator of effectiveness of secondary health care; and 3) to establish and compare the cost-utility of three large-volume surgical procedures in a real-world setting in the Helsinki University Central Hospital, a large referral hospital providing secondary and tertiary health-care services for a population of approximately 1.4 million.

**Patients and methods:** So as to identify studies that have used QALYs as an outcome measure, a systematic search of the literature was performed using the Medline, Embase, CINAHL, SCI and Cochrane Library electronic databases. Initial screening of the identified articles involved two reviewers independently reading the abstracts; the full-text articles were also evaluated independently by two reviewers, with a third reviewer used in cases where the two reviewers could not agree a consensus on which articles should be included.

The feasibility of routinely evaluating the cost-effectiveness of secondary health care was tested by setting up a system for collecting HRQoL data on approximately 4 900 patients' HRQoL before and after operative treatments performed in the hospital. The HRQoL data used as an indicator of treatment effectiveness was combined with diagnostic and financial indicators routinely collected in the hospital.

To compare the cost-effectiveness of three surgical interventions, 712 patients admitted for routine operative treatment completed the 15D HRQoL questionnaire before and also 3–12 months after the operation. QALYs were calculated using the obtained utility data and expected remaining life years of the patients. Direct hospital costs were obtained from the clinical patient administration database of the hospital and a cost-utility analysis was performed from the perspective of the provider of secondary health care services.

**Main results:** The systematic review (Study I) showed that although QALYs gained are considered an important measure of the effectiveness of health care, the number of studies in which QALYs are based on actual measurements of patients' HRQoL

is still fairly limited. Of the reviewed full-text articles, only 70 reported QALYs based on actual before–after measurements using a valid HRQoL instrument.

Collection of simple cost-effectiveness data in secondary health care is feasible and could easily be expanded and performed on a routine basis (Study II). It allows meaningful comparisons between various treatments and provides a means for allocating limited health care resources.

The cost per QALY gained was €2 770 for cervical operations and €1 740 for lumbar operations. In cases where surgery was delayed the cost per QALY was doubled (Study III).

The cost per QALY ranges between subgroups in cataract surgery (Study IV). The cost per QALY gained was €5 130 for patients having both eyes operated on and €8 210 for patients with only one eye operated on during the 6-month follow-up. In patients whose first eye had been operated on previous to the study period, the mean HRQoL deteriorated after surgery, thus precluding the establishment of the cost per QALY.

In arthroplasty patients (Study V) the mean cost per QALY gained in a one-year period was €6 710 for primary hip replacement, €52 270 for revision hip replacement, and €14 000 for primary knee replacement.

**Conclusions:** Although the importance of cost-utility analyses has during recent years been stressed, there are only a limited number of studies in which the evaluation is based on patients' own assessment of the treatment effectiveness. Most of the cost-effectiveness and cost-utility analyses are based on modeling that employs expert opinion regarding the outcome of treatment, not on patient-derived assessments.

Routine collection of effectiveness information from patients entering treatment in secondary health care turned out to be easy enough and did not, for instance, require additional personnel on the wards in which the study was executed. The mean patient response rate was more than 70 %, suggesting that patients were happy to participate and appreciated the fact that the hospital showed an interest in their well-being even after the actual treatment episode had ended.

Spinal surgery leads to a statistically significant and clinically important improvement in HRQoL. The cost per QALY gained was reasonable, at less than half of that observed for instance for hip replacement surgery. However, prolonged waiting for an operation approximately doubled the cost per QALY gained from the surgical intervention.

The mean utility gain following routine cataract surgery in a real world setting was relatively small and confined mostly to patients who had had both eyes operated on. The cost of cataract surgery per QALY gained was higher than previously reported and was associated with considerable degree of uncertainty.

Hip and knee replacement both improve HRQoL. The cost per QALY gained from knee replacement is two-fold compared to hip replacement.

Cost-utility results from the three studied specialties showed that there is great variation in the cost-utility of surgical interventions performed in a real-world setting even when only common, widely accepted interventions are considered. However, the cost per QALY of all the studied interventions, except for revision hip arthroplasty, was well below €50 000, this figure being sometimes cited in the literature as a threshold level for the cost-effectiveness of an intervention.

Based on the present study it may be concluded that routine evaluation of the cost-utility of secondary health care is feasible and produces information essential for a rational and balanced allocation of scarce health care resources.

Key words: cataract, cervical spine, cost-effectiveness, cost-utility analysis, hip replacement, health-related quality of life, knee replacement, lumbar spine, quality-adjusted life year, secondary health care

## Summary in Finnish

Räsänen Pirjo. Routine Measurement of Health-Related Quality of Life in Assessing Cost-effectiveness in Secondary Health Care [Terveysteen liittyvä elämänlaatu erikoissairaanhoidon kustannusvaikuttavuuden arvioinnissa]. STAKES, Research Report 163. Helsinki, Finland 2007. ISBN 978-951-33-1981-6

Järjestelmällisesti kerätty tieto hoidon vaikuttavuudesta ja kustannusvaikuttavuudesta on keskeisessä asemassa, kun erikoissairaanhoidon palveluja järjestetään tulevaisuudessa. Rajalliset resurssit ja terveysteknologioiden nopea kehittyminen edellyttävät toimintojen arviointia ja priorisointia. Päätösten tulee olla läpinäkyviä ja perustua tutkittuun tietoon, mutta ennen kaikkea myös siihen, miten potilaat itse kokevat hyötyvänsä erikoissairaanhoidon toiminnoista.

Tämän väitöskirjan tavoitteena oli kartoittaa tutkimukset, joissa hoidon lopputulosta arvioitiin laaturaportteina elinvuosina siten, että hoidon arvioijana oli potilas itse. Lisäksi väitöskirjan tavoitteena oli tutkia, miten erikoissairaanhoidon kustannusvaikuttavuuden arviointia voitaisiin toteuttaa Helsingin ja Uudenmaan sairaanhoitopiirissä. Erikoissairaanhoidon kustannusvaikuttavuuden arvioimiseksi tutkimukseen valittiin kolme lääketieteen erikoisalaa, jotka edustavat volyymiltaan suuria erikoissairaanhoidon ryhmiä, ja joilla siten on vaikutusta sekä kansantalouteen että kansanterveyteen.

Taloudellisten arviointien määrä terveydenhuollossa on kasvanut viimeisen kahden vuosikymmenen ajan. Tästä huolimatta kirjallisuuskatsauksemme perusteella löytyi vain vähän tutkimuksia, joissa hoidon lopputulosmittarina ovat laaturaportteiset elinvuodet, QALYt (Quality-adjusted life years) siten, että arvio hoidon vaikuttavuudesta on kysytty validilla menetelmällä potilaalta itseltään.

Hoidon vaikuttavuutta mitattiin 712 kirurgisesti hoidetun potilaan antamalla arviolla terveyteen liittyvästä elämänladusta ennen hoitoa ja noin 3–12 kuukauden kuluttua hoidosta. Terveysteen liittyvän elämänlaadun mittarina oli Suomessa kehitetty geneerinen, potilaan itsensä täyttämä 15D- kyselylomake. Kaikki potilaat saivat erikoissairaanhoidon normaalikäytännön mukaista hoitoa. Vaikuttavuustieto yhdistettiin hoidon kustannuksia ja järjestämistä koskevaan tietoon hoidon kustannusvaikuttavuuden arvioimiseksi.

Vaikuttavuustiedon rutiininomainen kerääminen erikoissairaanhoidon tulevilta potilailta osoittautui kohtuullisen helpoksi eikä vaatinut esimerkiksi lisähenkilöstön palkkaamista osastoille, joilla tutkimus toteutettiin. Vastusprosentti oli keskimäärin yli 70 prosenttia, mikä osoittaa sen, että potilaat osallistuvat mielellään tutkimukseen ja arvostavat sitä, että heidän voinnistaan ollaan kiinnostuneita vielä hoidon päättymisen jälkeenkin.

Tutkimustulosten mukaan sekä polvi- ja lonkkaproteesipotilaiden että lanne- ja kaularangan ahtaumapotilaiden terveyteen liittyvä elämänlaatu paranee seuranta-aikana merkitsevästi. Lonkan uusintaleikkaus- ja kaihileikkauspotilaiden elämänlaadussa paranemista tapahtuu vain vähän. Erikoissairaanhoidolle aiheutuneet kustannukset potilasta kohden vaihtelivat 1 640 € (kaihikirurgia) ja 11 240 € (lonkan uusintaproteesileikkaus) välillä. Kustannukset laatuainotteista elinvuotta kohden olivat halvimmat lannerangan ahtaumapotilailla, 1 740 €, ja kalleimmat lonkan uusintaproteesileikkauksessa, 52 270 €.

Johtopäätöksenä voidaan todeta että järjestelmällinen ja rutiininomaisesti toteutettu vaikuttavuustiedon kerääminen ja yhdistäminen kustannustietoihin on mahdollista melko vähäisin ponnistuksin. Tutkimus osoitti, että eri potilasryhmien kustannus laatuainotteista elinvuotta kohden vaihtelee suuresti, mutta on kaikissa muissa ryhmissä paitsi lonkan uusintaproteesileikkauksessa alle 50 000 €, jota alan kirjallisuudessa pidetään eräänlaisena kustannusvaikuttavan hoidon kattohintana. Laatuainotteiset elinvuodet ovat tärkeä terveydenhuollon vaikuttavuuden mittari. Arviointien, joissa laatuainotteiset elinvuodet perustuvat potilaiden itsensä antamaan arvioon saamastaan hoidosta ennen–jälkeen–asetelmassa, tulisi olla yksi keskeinen menetelmä pyrittäessä kustannusvaikuttavaan erikoissairaanhoidon.

Avainsanat: kaihi, kaularanka, kustannusvaikuttavuus, kustannus-utiliteettianalyysi, lonkkaproteesi, terveyteen liittyvä elämänlaatu, polviproteesi, lanneselkä, laatuainotteinen elinvuosi, erikoissairaanhoido



## Summary in Swedish

Räsänen Pirjo. Routine Measurement of Health-Related Quality of Life in Assessing Cost-effectiveness in Secondary Health Care [Rutinbaserad mätning av hälsorelaterad livskvalitet vid utvärdering av kostnadseffektivitet inom den specialiserade sjukvården]. STAKES, Research Report 163. Helsinki, Finland 2007. ISBN 978-951-33-1981-6

Systematiskt insamlad information om behandlingseffekt och kostnadseffektivitet är av central betydelse vid organiseringen av den specialiserade sjukvården i framtiden. De begränsade resurserna och den snabba utvecklingen inom olika hälsoteknologier förutsätter en utvärdering och prioritering av insatserna. Besluten skall vara transparenta och basera sig på undersökta fakta men framför allt på hur patienterna själva upplever sig ha nytta av insatserna inom den specialiserade sjukvården.

Syftet med denna avhandling är att kartlägga de undersökningar där behandlingens utfall uppskattats i kvalitetsjusterade levnadsår på så sätt att behandlingen bedömts av patienten själv. Dessutom avser avhandlingen att undersöka hur en utvärdering av den specialiserade sjukvårdens kostnadseffektivitet kunde genomföras inom Helsingfors och Nylands sjukvårdsdistrikt. För utvärderingen av den specialiserade sjukvårdens kostnadseffektivitet valdes till undersökningen tre medicinska specialområden, vilka till sin volym representerar stora grupper inom den specialiserade sjukvården och således påverkar både samhällsekonomin och folkhälsan.

Antalet ekonomiska utvärderingar inom hälso- och sjukvården har ökat under de senaste två årtiondena. Trots detta hittade vi på basis av vår litteraturöversikt endast få undersökningar, där kvalitetsjusterade levnadsår, QALY (Quality-adjusted life years) använts som behandlingens utfallsmätare på så sätt att man med en tillförlitlig metod bett patienterna själva ge en bedömning av behandlingseffekten.

Behandlingseffekten mättes utifrån en bedömning av den hälsorelaterade livskvaliteten hos 712 kirurgiskt behandlade patienter före behandlingen och cirka 3–12 månader efter behandlingen. Som mätare på den hälsorelaterade livskvaliteten användes ett i Finland utvecklat generiskt 15D-frågeformulär, som patienten själv fyller i. Alla patienter behandlades enligt normal praxis inom den specialiserade sjukvården. Uppgifter om behandlingseffekt kombinerades med uppgifter om behandlingens kostnader och arrangemang i syfte att utvärdera behandlingens kostnadseffektivitet.

Det rutinmässiga insamlandet av uppgifter om behandlingseffekten av de patienter som kom till den specialiserade sjukvården visade sig vara relativt lätt och krävde exempelvis inte avlöning av extra personal till de avdelningar där undersök-

ningen gjordes. Svarsprocenten var i genomsnitt över 70 %, vilket visar att patienterna gärna deltog i undersökningen och uppskattar att man visar intresse för hur de mår också efter att behandlingen slutförts.

Enligt undersökningsresultaten förbättras den hälsorelaterade livskvaliteten hos både knä- och höftprotespatienter samt hos patienter med förträngningar i länd- och halskotpelaren betydligt under uppföljningstiden. Vid omoperationer av höftleden och starroperationer sker endast en ringa förbättring av patienternas livskvalitet. Kostnaderna för den specialiserade sjukvården per patient varierade mellan 1 640 euro (starrkirurgi) och 11 240 euro (omoperation av höftprotes). Kostnaderna per kvalitetsjusterade levnadsår var lägst hos patienter med förträngningar i ländkotpelaren, 1 740 euro, och högst vid omoperation av höftprotes, 52 270 euro.

Som slutsats kan man konstatera att en systematisk och rutinmässig insamling av uppgifter om behandlingseffekten och en kombinerad av dessa uppgifter med kostnadsuppgifterna kan genomföras med relativt små ansträngningar. Undersökningen visade att kostnaderna per kvalitetsjusterat levnadsår varierar mycket mellan de olika patientgrupperna, men att kostnaderna i alla andra grupper utom omoperation av höftprotes är mindre än 50 000 euro, vilket inom litteraturen anses som ett slags takpris för en kostnadseffektiv behandling. De kvalitetsjusterade levnadsåren är en viktig mätare på hälso- och sjukvårdens effektivitet. Utvärderingar, där de kvalitetsjusterade levnadsåren baserar sig på patienternas egen bedömning av behandlingen i konstellationen före–efter, borde vara en viktig metod när man eftersträvar att uppnå en kostnadseffektiv specialiserad sjukvård.

Nyckelord: katarakt, halskotpelare, kostnadseffektivitet, kostnadsnyttoanalys, höftprotes, hälsorelaterad livskvalitet, knäprotes, ländkotpelare, kvalitetsjusterat levnadsår, specialiserad sjukvård

## Table of contents

Abstract	
Summary in Finnish	
Summary in Swedish	
List of original publications .....	15
List of abbreviations.....	16
1 INTRODUCTION.....	17
2 REVIEW OF THE LITERATURE.....	20
2.1 Economic evaluation .....	20
2.1.1 Why economic evaluation in health care? .....	20
2.1.2 Different types of economic evaluation.....	21
2.1.2.1 Cost-minimisation analysis .....	21
2.1.2.2 Cost-effectiveness analysis.....	21
2.1.2.3 Cost-utility analysis.....	22
2.1.2.4 Cost-benefit analysis .....	22
2.2 Costs.....	23
2.2.1 Costing in practice .....	24
2.3 Decision rules of cost-utility analysis.....	25
2.4 Quality-adjusted life year as an outcome measure.....	26
2.5 Measurement of health-related quality of life .....	28
2.5.1 Disease-specific instruments .....	29
2.5.2 Generic instruments .....	30
2.5.3 15D.....	30
2.6 Prior experiences of routinely evaluating cost-effectiveness in secondary health care.....	31
3 AIMS OF THE PRESENT STUDY .....	33
4 MATERIALS AND METHODS.....	34
4.1 Systematic review of the literature .....	34
4.1.1 Search strategy for the review.....	34
4.1.2 Screening of articles.....	35
4.1.3 Inclusion criteria .....	35
4.1.4 Evaluating the quality of the articles.....	35
4.1.5 Reporting.....	35
4.2 Feasibility study.....	36
4.3 Patients for the cost-utility analyses (III, IV and V).....	37
4.3.1 Neurosurgery (III) .....	37
4.3.2 Ophthalmology (IV).....	38
4.3.3 Orthopedic surgery (V).....	39
4.4 Effectiveness of the interventions.....	40
4.5 Costs.....	40
4.6 Statistics.....	41
4.7 Ethical considerations.....	42

5	RESULTS.....	43
5.1	Systematic review of the literature .....	43
5.2	Feasibility.....	44
5.3	Neurosurgery.....	45
5.4	Cataract surgery .....	45
5.5	Hip and knee replacement surgery .....	46
5.6	Costs of interventions .....	46
5.7	Quality-adjusted life years gained by the studied interventions.....	47
5.8	Cost per quality-adjusted life year.....	48
5.9	Sensitivity analyses.....	49
6	DISCUSSION .....	51
6.1	Limitations of the study.....	54
6.2	Future perspectives .....	56
7	CONCLUSIONS.....	57
	Acknowledgements .....	58
	References.....	60
	Original articles	

## List of original publications

This thesis is based on the following original studies, referred to in the text by their Roman numerals I–V

- I Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryyänen O-P, Roine RP. 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.
- II Räsänen P, Sintonen H, Ryyänen OP, Blom M, Semberg-Konttinen V, Roine R. Measuring cost-effectiveness of secondary health care: feasibility and potential utilization of results. *Int J Technol Assess Health Care* 2005;21:22–31.
- III Räsänen P, Öhman J, Sintonen H, Ryyänen O-P, Koivisto A-M, Blom M, Roine RP. Cost-utility analysis of routine neurosurgical spinal surgery. *J Neurosurg Spine* 2006;5:204–209.
- IV Räsänen P, Krootila K, Sintonen H, Leivo T, Koivisto A-M, Ryyänen O-P, Blom M, Roine RP. Cost-utility of routine cataract surgery. *Health Qual Life Outcomes* 2006;4:74. <http://www.hqlo.com/content/4/1/74>
- V Räsänen P, Paavolainen P, Sintonen H, Koivisto A-M, Blom M, Ryyänen O-P, Roine RP. Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthopaedica* 2007;78:108–115.

## List of abbreviations

15D:	15D HRQoL instrument
AQoL:	Assessment of Quality of Life instrument
CBA:	Cost-benefit analysis
CEA:	Cost-effectiveness analysis
CI:	Confidence interval
CINALH:	Cumulative Index to Nursing and Allied Health
CMA:	Cost-minimisation analysis
CUA:	Cost-utility analysis
DALY:	Disability-Adjusted Life Years
€:	Euro
EQ-5D:	EuroQoL HRQoL instrument
HHS:	Harris Hip Score
HRQoL:	Health-Related Quality of Life
HUCH:	Helsinki University Central Hospital
HUI:	Health Utilities Index HRQoL instrument (same as HUI3, HUI-Mark III)
HUS:	The Hospital District of Helsinki and Uusimaa
HYE:	Healthy Years Equivalent
ICER:	Incremental cost-effectiveness ratio
KSS:	Knee Society Score
LogMAR:	Logarithm of minimum angle of resolution
LYG:	Life Years Gained
NHS:	The National Health Service
NICE:	National Institute for Health and Clinical Excellence
N.S.:	Statistically not significant
QoL:	Quality of life
QALY:	Quality-adjusted life year
QWB:	Quality of Well-Being Scale
Rand-36:	The RAND 36-Item Health Survey 1.0
Rosser-Kind:	Rosser-Kind HRQoL instrument (same as The Charing Cross health indicator and Rosser index)
RS:	Rating Scale
RCT:	Randomized controlled trial
SAVE:	Saved-Young-Life Equivalent
SCI:	The Science Citation Index
SD:	Standard deviation
SF-36:	Medical Outcomes Study (MOS) 36-Item Short Form
SF-6D:	SF-6D HRQoL instrument
SG:	Standard Gamble
THA:	Total hip arthroplasty
TKA:	Total knee arthroplasty
TTO:	Time Trade-Off
VAS:	Visual Analogue Scale
VF-14:	Visual Function Index
Vs:	Versus

# 1 INTRODUCTION

All health care systems feel the strain of relentless economic scarcity. Strategies to tackle the emerging problems typically aim at increasing the effectiveness of the care delivery system, either by improving its allocative efficiency (whether to allocate resources to different groups of patients with different health problems) or its technical efficiency (resources are distributed among the same patient population in an attempt to maximise health gains within the budget available), though preferably both (Palmer & Torgerson 1999; Dowie 2001; Shiell et al. 2002; Cox et al. 2006). In an ideal health care system that produces maximal health benefits, the marginal cost-effectiveness of care is balanced, i.e., equal across all care programs or technologies used.

Traditionally, investments in health care have been made without detailed comparative information on the health gains attained. This can be considered irresponsible given that the allocation of scarce resources should be based on cost-effectiveness data, thereby ensuring the best possible return on investment. To reach this invaluable goal both the marginal effectiveness and the marginal costs tied into different technologies should be known. In reality, we usually do not know either of them.

Allocation decisions based on clinical results alone may lead to an 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 should, under optimal conditions, generate maximal benefits for the society but, especially in health care, allocation decisions are often combined with significant uncertainty.

In publicly funded health care systems, such as with the majority of specialised health care in Finland, limited resources mean that every available intervention cannot be provided in every situation to all who may need or want it. Choices must be made among effective health care interventions, and the decision to fund one can mean that others cannot be funded.

To estimate the allocative efficiency of care is the most important goal of all health care technology assessment. For this purpose, it often uses information derived from randomised controlled trials (RCTs). A problem with this approach, however, relates to the fact that the efficacy (the extent to which an intervention is active when studied under controlled research conditions) of any given technology thus measured does not nearly equate with that of its effectiveness (the extent to which an intervention produces an overall health benefit in routine practice) (Guide to Methods of Technology Appraisal, 2004). In most cases the cost-effectiveness of any technology may be severely reduced when the RCT study setting is replaced by

an everyday clinical environment. This is as a consequence of a systematic threefold bias, all in favour of the RCTs: 1) due to the multiple exclusion criteria common in RCTs, the patients lack the required similarity to the unselected clinical patients 2) the institutions participating in most trials represent the highest quality of any care under study instead of average providers; and, 3) in RCTs that consider surgical technologies, even the purely technical skills of the surgeons are systematically well above the average. Since all three factors overestimate the cost-effectiveness of any given technology, the resultant estimate may have little to do with the everyday cost-effectiveness of the technology (Hay 1999). Therefore, the effectiveness of any care cannot be extrapolated from its demonstrated efficacy. Instead, its true effectiveness must be measured.

The measurement problems, however, are not limited to the efficacy–effectiveness divide described above. Effectiveness, according to one often used interpretation, deals with so-called natural units of benefits gained, such as mortality, life-years, or days of absenteeism. According to this definition, the subjectively perceived quality of life does not enter the picture at all. Yet the everyday practice reveals that severely ill patients trade the quality of life actively even against the length of their own lives. Thus, without the inclusion of the health-related quality of life (HRQoL) aspect into the outcomes of all care, the rational medical decision-making may be distorted. Consequently, the way forward is to create a data collection system to measure the HRQoL of patients prior to and after most or all care delivery. Although of crucial practical significance, this has thus far not been systematically done anywhere in the world. In addition, also information about the costs related to outcomes is likewise needed to be able to estimate the cost-effectiveness of care as well.

After understanding the significance of effectiveness and HRQoL, a choice must be made between different HRQoL approaches, i.e., whether to use disease-specific or general HRQoL measures. The former are tailored to specific diseases, while the latter are constructed – as the name implies – for any given condition with reduced health. The disease-specific measures are more sensitive when measuring the point estimate or any change in any given condition. However, since the measurement is tuned according to the clinical expression of the measured condition, there is no way to compare the results in two different patient groups, not even within a single clinical specialty (like arthropathy and back-pain patients in orthopaedics). To make the outcomes measurement comparable, a generic HRQoL measure is thus necessary instead of a disease- or condition-specific HRQoL measure. The measurement of HRQoL by using a generic instrument that produces a single index number on a 0-1 scale –which reflects the trade-off between length and quality of life – also offers the significant advantage that effectiveness can be expressed in the form of quality-adjusted life years (QALYs) gained, which combine the change both in the length and quality of life. In recent years the QALY has been recognised as the most important indicator of effectiveness of health care interventions to



date. For instance, the National Institute for Health and Clinical Excellence (NICE), which provides national guidance on treatments and care for those using the NHS in England and Wales, uses the QALY as its principal measure of health outcomes (Rawlins & Culyer 2004; Rawlins & Dillon 2005).

In secondary health care, for instance in the framework of a large referral hospital like the Helsinki University Central Hospital (HUCH), the information on allocative efficiency of care is especially crucial. In a large organisation it is difficult, if not impossible, without systematically and routinely collected data to judge which intervention to use, what works, how well it works, at what cost, for whom, in what circumstances, and what is effective. In Finland, the idea of measuring the cost-effectiveness of secondary health care by using a standard generic HRQoL instrument (15D) was presented in 1996 (Sintonen). Pilot projects since have demonstrated that such an approach is feasible (Blom-Lange 1998; Kukkonen 2005).

Although previous smaller studies have indicated that systematic collection of HRQoL data is feasible, the question about the feasibility of large-scale routine data collection in the everyday practice of a large hospital, i.e., not in the framework of a research project, still remains open. Therefore, the aim of this study was to test the feasibility of collecting effectiveness data using the generic 15D HRQoL instrument in several medical units in the HUCH, and of combining these data with routinely collected cost data to create a means for estimating the cost-effectiveness of secondary and tertiary health-care services.

## 2 REVIEW OF THE LITERATURE

### 2.1 Economic evaluation

An economic evaluation is a comparative analysis of alternative courses of action in terms of both their costs and consequences (Drummond et al. 2005). With economic evaluation it is possible to get information about whether an intervention is cost-effective, i.e., worth providing or not (Guidelines for the Economic Evaluation of Health Technologies, 2006). The basic tasks of any economic evaluation are to identify, measure, value, and compare the costs and consequences among the alternatives being considered (Drummond et al. 2005).

Central to the area of health economics are the concepts of opportunity cost and incremental change. Relevant information is generated by economic evaluation, which measures the relation between the health effects and costs of various interventions. The basic question in economic evaluation centres on incremental cost-effectiveness: On the one hand, what are the health effects of an intervention that is costed in monetary units, and on the other, what health benefits are lost when the resources are not invested in the alternative intervention.

Appropriate criteria that should be used in setting priorities in a publicly funded health care system are as yet unfortunately far from clear. From a health economic perspective, one criterion that might be considered as part of the decision-making process when setting health care priorities is the maximisation of QALYs (Stirling et al. 2002).

#### 2.1.1 Why economic evaluation in health care?

In health care choices must be made between services, as scarce resources preclude the use of all available interventions. The importance of right choices becomes more and more evident as the population is ageing and the variety of new, usually expensive, interventions seems to be ever increasing.

The fundamental purpose of health care is to produce as much health as possible. This means that the resources available should be directed to interventions that are as effective as possible in relation to their costs. In order to ensure that health care resources are used in the most appropriate manner and to achieve maximal patient benefit, health care decision-makers need to adopt robust processes for setting priorities. For this they need economic evaluations that provide information that is useful, relevant, and timely (Raftery 1998; Drummond & McGuire 2001). Unfortunately, the choices made in health care are not always based on scientific evidence on the effectiveness, let alone the cost-effectiveness, of interventions.

Effective health care is achieved in an ethically sound way by choosing the most cost-effective interventions. Efficient health care makes it possible to offer patients the best possible care by the means of the most cost-effective interventions. However, the maximisation of efficiency, i.e., cost-effectiveness, must not be such an overriding goal that other considerations such as equity are forgotten (Drummond et al. 2005; Tugwell et al. 2006). Deviations from efficiency maximisation are desirable, indeed necessary, on the basis of commonly accepted, and preferably officially endorsed, equity criteria, which reflect the value climate in the society.

## 2.1.2 Different types of economic evaluation

There are four main types of economic evaluation: cost-minimisation analysis (CMA), cost-effectiveness analysis (CEA), cost-utility analysis (CUA) and cost-benefit analysis (CBA). The difference between them relates to how the health effects of health care interventions are measured and valued. The purpose of the study and what kind of data is available are crucial when choosing the type of the analysis.

### 2.1.2.1 Cost-minimisation analysis (CMA)

The CMA is applicable if the effectiveness of the interventions being compared has been shown to be the same. Then only the cost difference between the interventions is determined. A CMA asks the question, “Which intervention is the least expensive given that they are equally effective?” As the clinical effectiveness of the interventions being compared is only rarely exactly the same, the usefulness of CMA is limited. It can be employed, for instance, in cases where a new treatment shortens hospital length of stay without otherwise affecting the cure of the patient compared to the alternative treatment.

### 2.1.2.2 Cost-effectiveness analysis (CEA)

In CEA it is necessary to identify, measure, and value both costs and health outcomes. Health outcomes are measured in simple, one-dimensional natural units, for example life years gained or disability avoided. CEA reveals the relationship between the resources used (costs) and the health benefits achieved (effects) for a health or medical intervention compared with an alternative strategy. CEA asks the question, “What is the comparative cost of the two interventions per outcome?” (Petitti 2000). A CEA can also utilise structured health or quality of life measures that have not been valued with regard to population preferences.

### 2.1.2.3 Cost-utility analysis (CUA)

CUA is very similar to CEA, but the outcomes used for CUA differ from those of CEA. In fact, CUA is a form of CEA, in which the cost of the studied intervention is related to the change in the length and/or quality of life resulting from the intervention. It thus allows for HRQoL adjustments to a given set of treatment outcomes, while simultaneously providing a generic outcome measure for a comparison of costs and outcomes in different programs. The health outcome is usually measured as the number of QALYs gained. The results of CUAs are typically expressed in terms of cost per QALY gained. Despite differences in the outcomes employed, the terms CUA and CEA are often used as synonyms (Drummond et al. 2005). Even if utility is an economic concept that reflects the desirability or value of an activity to an individual, a cost-QALY analysis would be a more illustrative term for this type of analysis than cost-utility analysis.

CUA can be recommended especially for situations in which HRQoL can be considered an important outcome of the studied intervention. It also serves as an advisable way to analyse outcomes when the studied intervention affects both morbidity and mortality, or the intervention has a wide range of different outcomes and it is necessary to have a common unit that takes into account all of them. CUA may also help decision-makers to set priorities and to decide which interventions to downgrade or eliminate when facing limited financial resources (Drummond et al. 2005).

A CUA, however, is sometimes problematic. The different methods and instruments that are used for measuring preferences for health states often produce different scores for the same health state. Furthermore, comparing a chronic condition with an acute or a mild to severe condition may pose difficulties. Consequently, the results of a CUA – as with the results of other types of studies – must always be interpreted with care.

Although it is not without its critics, cost-utility analysis is recommended for several reasons: It captures gains from both prolongation and quality of life in a single measure, it incorporates the value that people place on different health outcomes or their preferences for particulate health states, and it provides a convenient means of comparing analyses of diverse interventions and conditions (Neumann et al. 2000).

### 2.1.2.4 Cost-benefit analysis (CBA)

In CBA the costs and health consequences of treatment are measured and valued in monetary units. Consequently, CBA relates the net monetary gain of treatment to its costs. CBA addresses the question, “what is the overall economic trade-off between the interventions?” The results of CBA may be stated either in the form of a ratio of costs to benefits, or as their simple sums referred to as net benefit

(Drummond et al. 2005). The strength of CBA is that it may be easier to understand than analyses employing QALYs or life-years gained (LYG). On the other hand, the monetary value of a health gain is in most instances difficult, if not impossible, to determine. Although CBA may be useful in some situations, it should generally be regarded as a secondary analysis.

## 2.2 Costs

It is usually recommended to carry out economic evaluations from a societal perspective, which entails identifying, measuring and costing the use of all resources related to the alternatives being compared, regardless of to whom the costs accrue. The resource use and associated costs are classified into several categories.

Direct health care costs include the value of health care resources directly consumed by the interventions to be evaluated: capital goods like building site, buildings, facilities, equipment, energy, materials and consumables like hospital and other prescribed drugs, and working time of different types of personnel. The value of resources used in treating any possible side effects of the intervention also belongs to direct health care costs.

Direct non-health care costs include the value of resources from outside health care that are directly required when using the service under evaluation. Such resources are, for instance, social services like home help or care in social sector institutions, special schooling, child care and voluntary services, over-the-counter medicines, special clothing and diet, appliances, home adaptation, and travel costs. Direct non-health care costs also encompass the value of time directly required from the patients when they use the service, including travel and waiting time, time required by diagnostics and treatment, and the value of time required from the carers (parents, spouse, relatives etc.) at home and when accompanying the patient to health service facilities. The value of time required from the patients and carers is also referred to as time costs (Sintonen 2007). The value of healthy time should also be considered when assessing non-health care costs, although this does pose difficulties (Drummond et al. 2005). Productivity costs, earlier referred to as indirect costs, arise from reduced productivity due to illness affecting time at work or complete temporary (sick leave) or permanent loss of ability to work due to disability and/or premature death. The threshold at which time costs turn into productivity costs is vague.

The future health care costs (or savings) that are directly attributable to the intervention are the increased (or decreased) costs due to a health care intervention's effect on the individual's longevity. The handling of unrelated future medical costs is important because they can be large enough to raise the CE ratio substantially. The impact is greatest when the intervention primarily extends life (Garber & Phelps 1997). There is some controversy over how to treat the so-called future costs

(Stone et al. 2000; Petitti 2000). They should include the costs, benefits and cost-savings related to the disease being studied, but it is still under debate whether they should also include costs for diseases unrelated to the intervention in question (Gold 1996).

Which costs are included in the analysis will vary depending on the timeframe and perspective being considered in a study. Most guidelines suggest the use of a societal perspective (Drummond & Jefferson 1996; Gold et al. 1996). This means that also costs accrued outside health care should be included.

Stone et al. (2000) identified in a literature review the cost components that have been included in published CUAs. They found that in nearly all of the included studies the analysts had included some direct health care costs. Direct non-health care costs and time costs were generally lacking, with only 17 % including them, as were productivity costs, which were included in only 8 % of the studies.

## 2.2.1 Costing in practice

The costing procedure can be divided into three phases: identification, measurement and valuation.

First it is important to identify the relevant use of resources involved in the health technologies to be compared. The choice of perspective for the analysis is momentous. Methods such as a review of earlier studies in the area, pilot studies, modelling exercises or expert advice can help to identify the relevant resource use (Kristensen et al. 2001).

In measuring resource use the typical units are physical units like time consumption of health professionals, inpatient days, medicines taken, number of tests and operations, number and duration of visits to a general practitioner, days of absence from work, etc. Potential methods and data sources for the collection and measurement of resource use are case record forms, cost diaries, questionnaires or interviews, register and data files, pilot studies, clinical databases, earlier studies or expert advice. It is often a good idea to combine some of the methods (Kristensen et al. 2001).

The value of resource use should ideally be the opportunity cost. Opportunity costs, however, are difficult to measure in practice. Market prices are only one way of valuing the costs of resource use; the others are charges, average costs and marginal costs. When real costs are measured as part of a micro-costing or a patient-specific costing approach, either marginal or average costs might be used (Larsen et al. 2003). Marginal costs are the costs that are expected to vary according to the number of patients treated, and express the cost of using one extra unit of resources. Average costs also include the resource use that does not vary with the number of patients treated, like for instance capital costs and other fixed costs (Larsen et al. 2003).

Stone et al. (2000) evaluated sources of cost valuations in cost-utility analyses and found that the most frequently used source was published estimates, followed by charges, institutional-based accounting and expert opinion.

## 2.3 Decision rules of cost-utility analysis

To differentiate which health technology is cost-effective, both the total costs and the effectiveness of at least two interventions have to be compared. The comparison may lead to nine different situations described in the decision-matrix below (Table 1).

TABLE 1. The cost-effectiveness decision-matrix adapted from Drummond et al. 1997

A new technology compared with an old one	Less effective	Same effectiveness	More effective
Less costly	1.No clear decision non-dominance => Incremental analysis needed	4.Adopt the new technology the new dominates the old (weak dominance)	7.Adopt the new technology the new dominates the old (strong dominance)
Same costs	2.Keep the old technology the old dominates the new (weak dominance)	5.The technologies are equal	8.Adopt the new technology the new dominates the old (weak dominance)
More costly	3.Keep the old technology the old dominates the new (strong dominance)	6.Keep the old technology the old dominates the new (weak dominance)	9.No clear decision non-dominance => incremental analysis needed

In situations described in cells 1 and 9 marginal or incremental analysis is needed to decide which technology is preferable. For that purpose the incremental cost-effectiveness ratio (ICER) has to be calculated. It is a ratio of the difference in costs of interventions to the difference in outcomes. The new intervention is considered cost-effective if the society is willing to pay for the additional benefits (cell 9) or if the society considers that the cost savings compensate for the lower effectiveness (cell 1).

When calculating the ICER in the situation depicted in cell 9, one inevitably faces the question whether there is a willingness-to-pay threshold, which indicates what might be the maximum incremental cost that the society is willing to pay for

a QALY gained. Some countries use a formal threshold. Rawlins and Culyer (2004) suggest that NICE in the UK bases decisions primarily on an ICER below £20 000/QALY gained. However, Parkin and Devlin (2004) found that NICE's threshold is probably between £20 000–30 000/QALY gained. To determine a universally accepted, official threshold value for the willingness to pay for a QALY is an impossible task as resources available for health care vary from country to country, and also other factors like uncertainty and burden of disease must be taken into consideration (Devlin & Parkin 2004).

## 2.4 Quality-adjusted life year as an outcome measure

CUA has increasingly been recognised as the preferred method to assess the cost-effectiveness of care (Zentner et al. 2005). Following on from this preference, health outcomes can be measured in terms of QALYs gained, which is an outcome measure that takes into account and combines within a single measure both changes in the quantity and/or the quality of life produced by a health care intervention. It is an arithmetic product of the life expectancy and the quality of the remaining life years. QALYs are often used to describe health outcomes of interventions, based on data collected at baseline and at one or more points of follow-up (Manca et al. 2004). As QALYs can be used to express the sum of individual utility gains from medical interventions, they are, as such, potentially a useful measure in deciding how to allocate resources (Nord 1999).

The calculation of QALYs can be presented by the simple equation:  $QALYs = U_i * t$  (where  $U_i$  represents the utility score for health state  $i$  and  $t$  represents the amount of time spent in health state  $i$ ). The example in Figure 1 below shows how the length of life (horizontal axis) and the health outcome in terms of HRQoL (vertical axis; on a 0–1 scale) can be combined. The values on a 0–1 scale are referred to in the literature as utilities, preference weights, quality weights or values of health states. In the example the duration of health state 1 is  $k_1$  and its utility  $v_1$ . State 1 is followed by state 2, the duration of which is  $k_2$  and the utility  $v_2$ . Finally, state 3 with a utility of  $v_3$  lasts until death. Thus the sum  $k_1v_1+k_2v_2+k_3v_3$  (white area under the curve) depicts the number of QALYs without treatment ( $QALY_0$ ). With treatment HRQoL is improved to a higher level and then HRQoL deteriorates more slowly and the length of life is longer. The area under the upper curve represents the number of QALYs with treatment ( $k_4v_4+k_5v_5+k_6v_6+k_7v_7 = QALY_w$ ) and the difference between the two areas (grey area) the incremental number of QALYs gained by treatment ( $QALY_w - QALY_0$ ). This area is used as the indicator of effectiveness of treatment in CUA (Drummond et al. 2005; Sintonen 2007).





equal to a gain of 25 QALYs each to four individuals). (Guidelines for economic evaluation of pharmaceuticals, 1997)

The healthy years equivalent (HYE), the disability-adjusted life years (DALY), and the saved-young-life –equivalent are alternatives to the QALY (Drummond et al. 2005). The DALY, however, has been criticized for several of its properties (Arnesen & Nord 1999) and has not gained as widespread acceptance as the QALY. The same applies to HYE and SAVE: In a systematic literature review evaluating methods used in cost-utility studies between 1981–2000, Greenberg & Pliskin (2002) were able to identify only one study that used cost per HYE and none that used cost per SAVE.

To be able to express the effectiveness of treatment in terms of QALYs we need HRQoL measurements that describe the various dimensions of health and their preference weights elicited from patients or the general public.

## 2.5 Measurement of health-related quality of life

Quality of life (QoL) is a broad concept that includes many aspects of life in addition to health; for example wealth, freedom, the political system, and the cleanliness of the environment all contribute to the overall QoL. HRQoL refers to those aspects of QoL that are related to health (Guidelines for the Economic Evaluation of Health Technologies, 2006). Consequently, HRQoL is a more concise concept than quality of life per se and considers mainly aspects of life that can be affected by health care or health policies.

HRQoL reflects the physical, psychological and emotional dimensions of health (Gold et al. 1996). It 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 (Kauppinen 1999). 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 he may have, and his phase of life, as well as his activities and goals.

There are numerous ways to assess HRQoL: disease-specific, generic, or preference-based (utility) measurements tools. The different dimensions of health can be described with various generic or disease-specific instruments. The generic instruments allow the comparison of a variety of patient groups, whereas the disease-specific instruments only give information on the effect of a certain disease on health and are thus not suited for a comparison of treatment results across different diseases.

As already stated above, the terms utility, value and preference are often used as synonyms, although strictly speaking there are differences between them.

Drummond (2005) defines preference as an umbrella term that describes the overall concept; utilities and values are different types of preferences. Preference weights differ from measures of functional health status, because they measure how patients or population feel about their health and ability to function rather than how they actually function (Tosteson et al. 2000). Preference-based measures provide a summary score that numerically reflects the HRQoL, and are the only approaches that are suitable for use in a CUA.

The choice of utility instrument can be expected to have a large impact on the results of cost-utility studies (Stavem 1999). For example, Brazier et al. (2004) found in their study that when comparing seven patient groups, there are differences between the results obtained with different HRQoL measures (EQ-5D vs. SF-6D). Reasons for this are the differences in the health state descriptive system and the methods used to value the health states defined by the descriptive system. So far, however, no single measure of HRQoL has been accepted as the gold standard.

There are two major methodological problems in constructing an HRQoL measure that produces a single index score. One is the creation of a standardised health state descriptive system, i.e., the choice of the dimensions (attributes) in which health is to be measured and the division of each dimension into discrete levels, by which more or less of the attribute can be identified. Another is the valuation of the different combinations (profiles) of the levels, one from each dimension, i.e., of the health states thus defined.

### 2.5.1 Disease-specific instruments

Any generic instrument that is simple enough to be applied across different interventions may lack sensitivity to important differences in health status that are essential for particular diseases. For that reason disease-specific measurements have an important role in assessing the effectiveness of treatment (Petitti 2000). Disease-specific instruments focus on the issues most relevant to patients' health concerns, having the detail to detect small, though clinically important changes in health status. They are thus not suited for a comparison of treatment results across various diseases (Drummond et al. 2005). Their main purpose is to assist in clinical decision-making, though they are also usually sensitive in measuring the results of specific treatments. Good examples of disease-specific instruments are for instance the Knee Society Score (KSS), which evaluates pain and mobility in patients with knee problems (The Knee Society 2006), and the Harris Hip Score (HHS), which has been designed for the assessment of symptoms of hip disorders (Harris 1969).

## 2.5.2 Generic instruments

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 describe the health state from the stand point of various physical and emotional dimensions such as vitality, role limitations caused by emotional difficulties, bodily pain, general health, social function, etc. as in the widely used SF-36 instrument (Ware et al. 1995). The latter instrument produces 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 QALYs used for a commensurate appraisal of the cost-effectiveness of various health care interventions. When choosing an HRQoL instrument, special attention needs to be paid to its empirical, theoretical, and technical characteristics, such as validity, reliability, sensitivity, usability, and interpretability (Gerard 1992; Fitzpatrick et al. 1998; Drummond et al. 2005). Generic, single index score instruments include for instance the EQ-5D (EuroQol), SF-6D (derived from RAND-36/SF-36), HUI 3 (Health Utilities Index Mark III), the AQoL (Assessment of Quality of Life) and the 15D (Gerard et al. 1999; Drummond et al. 2005).

HRQoL instruments differ regarding the methods they use for measuring preferences and from whom the preferences are obtained. The same patient group can get a quite different HRQoL score on the scale of 0 to 1 depending on the method used. So far, there is no consensus which HRQoL instrument is the most effective to use.

## 2.5.3 15D

The development of the 15D questionnaire started in the late 1970s. The idea of the instrument was to combine the advantages of a profile and a preference-based measure. The first version in the early 1980s was a 12-dimensional system (Sintonen 1981). The first version of the 15-dimensional health state descriptive system emerged in 1986 (Sintonen & Pekurinen 1989 and 1993).

The present version of the 15D originating in 1992 consists of 15 dimensions (moving, seeing, hearing, breathing, sleeping, eating, speech, eliminating, usual activities, mental function, discomfort and symptoms, depression, distress, vitality and sexual activity) each with five levels, leading to  $5^{15}$  possible combinations of answers describing different health states. For each dimension, the respondent must choose one of the five levels that best describes his/her state of health at the moment (the best level = 1; the worst level = 5).

The valuation system of the 15D is based on an application of the multi-attribute utility theory. A set of utility or preference weights, elicited from

representative samples of the general public through a 3-stage valuation procedure, is used in an additive aggregation formula to generate the utility score, i.e., the 15D score (single index number) over all the dimensions. The maximum score is 1 (no problems on any dimension), and minimum score 0 (being dead). The scores are suitable for QALY calculations. A change in the score of  $\geq 0.03$  is considered clinically significant or important (Sintonen 1995).

In most of the important properties the 15D compares favourably with other instruments of that kind (Stavem 1999; Hawthorne et al. 2001; Sintonen 2001) and the instrument has been widely used in patients with different diseases in different countries. The questionnaire is available in more than 20 languages. Based on the 15D, a 16D version has also been developed for adolescents aged 12–15 years and 17D for children aged 8–11 (Apajasalo et al. 1996 a & b; [www.15D-instrument.net](http://www.15D-instrument.net)).

## 2.6 Prior experiences of routinely evaluating cost-effectiveness in secondary health care

In Finland, the idea of measuring the cost-effectiveness of secondary health care by using a standard generic HRQoL instrument (15D) was presented in 1996 (Sintonen 1996). Pilot projects since have demonstrated that it is feasible (Blom-Lange & Sintonen 1998).

Blom-Lange (1998) and Blom-Lange & Sintonen (1998 & 2000) evaluated the feasibility of measuring the effectiveness of routine hospital treatment in the Päijät-Häme Central Hospital, Finland. The main aim of the research was to define the most appropriate point of time to assess HRQoL. A total of 645 patients in eleven diagnostic groups filled in the 15D questionnaire at five different time-points: when placed on the waiting list, at admission to the hospital, at discharge from the hospital, approximately three to four months after discharge and one to two years thereafter. The results of the studies indicated that it is possible to use the 15D instrument to assess the effectiveness of hospital treatment. In all diagnostic groups the average 15D scores were lower at the final measurement than three or four months after the intervention. In addition, the research showed that patients' health status did not deteriorate while on the waiting list.

Kukkonen (2005) studied 828 patients in the North-Karelia Hospital District in Finland by administering the 15D HRQoL instrument to patients before and up to four months after treatment in specialised health care. The HRQoL instrument proved to be comprehensive enough to be used in a mixed patient population. It was also shown to provide valuable input for Bayesian mathematical modelling. It was concluded that in the area of the North Karelian hospital district in Finland, specialised health care can mostly be considered cost-effective.

A similar approach to the measurement of effectiveness has been used in Canada, where outcomes of elective surgery in a large group of surgical patients were successfully evaluated using the SF-36 (Wright et al. 2002). However, in the study of Wright et al. cost data were not used and, consequently, cost-effectiveness not reported.

### 3 AIMS OF THE PRESENT STUDY

The main aim of the present research was to explore the feasibility of the large-scale routine collection of effectiveness data using the generic 15D HRQoL instrument in the Hospital District of Helsinki and Uusimaa.

1. The first step of the study was to identify, in a systematic literature review, the cost-effectiveness studies that used QALY based on actual measurements of patients' health-related quality of life as an outcome measure (Study I).
2. Secondly, building on previous experiences, we wanted to establish a system for the large-scale collection of effectiveness data and to test the feasibility of such an approach in the context of a large university hospital (Study II).
3. The third aim was to combine HRQoL data with the available cost data to create a means for estimating the cost-effectiveness of secondary and tertiary health care services routinely (Study II).
4. The fourth goal was to report and compare the cost-utility of three different important surgical interventions routinely performed in the Hospital District of Helsinki and Uusimaa (spinal surgery, cataract surgery and hip and knee replacement surgery) (Studies III, IV, and V).

## 4 MATERIALS AND METHODS

This thesis is based on three types of studies. Article I is a systematic literature review of studies that have used QALY as an outcome measure. Article II assesses the feasibility of the routine collection of HRQoL data in secondary health care. In addition, article II explores methods of combining these effectiveness data with the diagnostic and financial indicators of care routinely collected in the HUCH. Articles III-V report on economic evaluations for three routinely offered surgical interventions that constitute a significant part of the activities in the medical specialties of neurosurgery, ophthalmology and orthopaedics.

### 4.1 Systematic review of the literature

A systematic review of the literature is a research article that identifies relevant studies, appraises their quality and summarises their results using scientific methodology (Khan et al. 2003). The systematic review addresses questions formulated through an analysis of the evidence. It has explicit methods for searching of the literature, as well as having critical inclusion and exclusion criteria. The aim is to ensure a review that is comprehensive and unbiased and which can be used with confidence for decision-making in the research and delivery of health care.

#### 4.1.1 Search strategy for the review

Systematic searches of the literature were made, 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) and the search strategies described in detail in Appendix A of the article. 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/>).



### 4.1.2 Screening of articles

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.

### 4.1.3 Inclusion criteria

The articles deemed acceptable for inclusion had compared in a scientifically valid manner the HRQoL of patients in a before–after setting and where patients had assessed their HRQoL with a generic HRQoL instrument recognised to produce a valid single index score for the calculation of QALYs (15D, EQ-5D, SF-6D, HUI, AQL, QWB, Rosser-Kind), or in which the HRQoL had otherwise been assessed by a direct valuation method (Time Trade-Off (TTO), Standard Gamble (SG), Visual Analogue Scale (VAS), or Rating Scale (RS)).

### 4.1.4 Evaluating the quality of the articles

The quality of the included articles was considered with regard to the study design used and study performance as reported by Hailey et al. (2004 a & b). For study design, scores were assigned to five classifications. Large randomised controlled trials (RCTs), defined as those with at least 50 subjects in each arm, were given a score of 5. Small RCTs had a score of 3, prospective non-randomised studies 2, retrospective comparative studies 1, and non-controlled series 0.

Study performance was established by looking at 5 aspects: patient selection, description/specification of the intervention, specification and analysis of study, patient disposal, and outcomes reported. When reviewing a study, each of these five aspects was given a score of 0, 1 or 2. A score of 0 was given if the relevant information was missing or given in only minimal detail; a score of 1 indicated that reasonable detail was provided, but there were some important limitations; and a score of 2 was given if the information provided was satisfactory, with no significant limitations. Each study therefore had a possible maximum score of 10 for performance.

### 4.1.5 Reporting

The included articles were summarised in a table with regard to country of origin, intervention, clinical specialty, aim of the study, subjects, methods, perspective of the analysis, HRQoL instrument used, costs, results, limitations and cost/QALY.

## 4.2 Feasibility study

Both the feasibility study and the studies forming the bases of articles III-V were carried out in secondary health care. Secondary health care includes treatments that are provided both on an outpatient and inpatient basis in hospitals. For the organisation of specialised health care, Finland is divided into 20 hospital districts. Five of them are university hospital districts. All of them are owned by federations of municipalities, i.e., hospital districts (Järvelin 2002). The Hospital District of Helsinki and Uusimaa (HUS) is the largest hospital district in Finland. HUCH is part of HUS and it is a large referral hospital that provides secondary and tertiary health-care services for a population of approximately 1.4 million ([www.hus.fi](http://www.hus.fi)). The Hospital District of Helsinki and Uusimaa has 32 hospitals throughout the province of Uusimaa. Specialised medical care in Finland refers to that part of health care that is run by specialists and which covers, in addition to diagnostics and treatment, some preventive and rehabilitative services. Access to specialised care usually requires a referral from a primary care physician except in emergency situations. All of the major medical specialties are represented at HUS: surgery, medicine, anaesthesiology, physiatrics, obstetrics and gynaecology, illnesses of children and adolescents, neurology, neurosurgery, ophthalmology, otorhinolaryngology, imaging, laboratory specialties, psychiatry, oncology, dermatology and allergology, and venereal diseases.

Data collection began in March 2002 and has so far covered over 11 000 patients receiving treatment according to normal hospital routines in over 30 distinct medical entities. HRQoL was measured before and after treatment by the 15D HRQoL questionnaire. Some disease-specific measures were also used for some medical entities.

Direct costs of secondary health care were obtained from the Ecomed® patient administration system (Datawell Ltd., Espoo, Finland) used in the HUCH. This system includes, in addition to cost data, data on patient variables such as diagnoses and interventions performed, and organisation of services such as waiting times and urgency of treatment. To evaluate the cost-effectiveness of treatments, the HRQoL data are combined with these data. The summary costs used for the analysis covered all relevant specialty-related costs including pre- and postoperative outpatient visits to the hospital. To ensure this, collection of cost data covered a period starting three months prior to the date on which the first HRQoL questionnaire was completed and ending at the date the repeat questionnaire was completed. For each treated patient, all available information on treatment in the database (for instance diagnosis, intervention, costs, waiting time) were combined for the above-mentioned time period with the HRQoL information.

## 4.3 Patients for the cost-utility analyses (III, IV and V)

### 4.3.1 Neurosurgery (III)

In addition to life-saving interventions (like for instance surgery of intracranial aneurysms) neurosurgeons often perform operations aimed mainly at relief of pain. These include operations for benign back disorders which, although they usually have a good long-term prognosis, may severely reduce a patient's HRQoL (Daffner et al. 2003) and be associated with prolonged suffering and inability to work.

Cervical disc disease accompanied by spondylosis may cause neurological dysfunction via ventral compression of the spinal cord or nerve roots, neck pain, or a combination of symptoms (Angevine et al. 2005). Many studies have shown that a variety of cervical interventions lead to positive postoperative results judged by both physician-assessed (Bertalanffy et al. 1988; Gaetani et al. 1995; Bucciero et al. 1998; Sampath et al. 2000; Goldberg et al. 2002) outcomes as well as patient-reported outcomes.

Cervical stenosis is a degenerative disease where the spinal canal and neural foramina narrow and compress the spinal cord and nerve roots. Low back (lumbar spine), mid back (thoracic spine) and neck (cervical spine) pain syndromes adversely affect the function or well being of an individual and are attributable to many non-malignant aetiologies. Lumbar radiculopathy is a common problem that results when nerve roots are compressed or irritated. Low back pain resulting from degenerative disease of the lumbosacral spine is a major cause of morbidity, disability and lost productivity. Treatment for symptomatic lumbar stenosis is usually surgical decompression.

In the HUCH neurosurgeons perform around 3000 operations annually. Of them approximately 700 (23 %) represent demanding back operations. The majority of the operations concern the cervical spine (over 400 operations annually), the rest complicated disorders of the thoracic or lumbar spine (<http://www.hus.fi>).

Between April 2002 and April 2003 546 patients undergoing spinal operations at the Department of Neurosurgery of the HUCH were asked to complete the 15D-questionnaire. Of them 70% returned the baseline questionnaire. The follow-up questionnaire was administered approximately three months after the operation and was returned by 311 patients (82 % of those that responded to the first survey). However, 12 patients were excluded from the final analysis because they had filled in the HRQoL survey after the operation and five patients because of incomplete data.

Altogether 270 patients were available for the final analysis. As the total number of spinal operations in the department of Neurosurgery in 2002 was 687,

our sample represents approximately 40 % of the total number of patients operated on for spinal disorders per year.

### 4.3.2 Ophthalmology (IV)

A cataract is opacity of the lens of the eye that disturbs the passage of light to the retina, thus impairing vision. Vision problems like cataract are conditions that may not directly affect the length of life but may negatively affect its quality.

According to the Finnish Current Care guideline a cataract operation is indicated if it impairs a patient's quality of life and ability to manage with work or daily activities and the patient wishes to have surgery after having received adequate information on the expected benefits and risk factors of surgery. Special indications for cataract surgery include the treatment and monitoring of diseases of the posterior eye (for instance diabetic retinopathy, retinal diseases), open angle glaucoma, phacodonesis and the repair of anisometropia. Cataract surgery is also indicated if a hypermature cataract causes an inflammation in the eye (phacolysis, phacoanaphylaxis), constriction of the iridocorneal angle or a refractory open-angle glaucoma (Käypä hoito -suositus [Current Care Guideline] [www.kaypahoito.fi](http://www.kaypahoito.fi)).

Cataract surgery is the most common surgical operation in ophthalmology. The rate of cataract surgery in Finland has been increasing every year, though since the recent introduction of national guidelines for treatment indications for cataract, the number of operations has shown a small decline (Punnonen 2007). In 2002, over 40 000 cataract operations were performed in Finland (Käypä hoito -suositus [Current Care Guideline] [www.kaypahoito.fi](http://www.kaypahoito.fi)). In the future the demand for cataract surgery is expected to increase as the population ages.

In the field of ophthalmology 386 patients scheduled for routine cataract operation in the Helsinki University Eye Hospital between August 2002 and June 2003 were invited to participate and complete the 15D questionnaire. A follow-up survey was mailed approximately six months after the operation to all patients who returned the baseline questionnaire. The follow-up questionnaire was returned by 73 % of those that responded to the first survey. Seven cases were removed from the final analysis because of incomplete data, 32 patients because they had filled in the baseline questionnaire after the operation, six patients because they had filled in the follow-up questionnaire less than 60 days after the operation on the second eye, seven patients because the eventual intervention performed was more extensive than a plain cataract operation, and 11 patients because their final principal diagnosis was not cataract. 219 patients were available for final analysis. In 2002, the total number of cataract operations in the HUCH was 4862, of which 1164 were performed in the ward used for our study. Our sample thus represents

approximately 19 % of the cataract operations performed per year in the ward taking part in the study.

### 4.3.3 Orthopaedic surgery (V)

In the field of orthopaedics hip and knee replacement comprise a major area. The number of hip and knee operations has been constantly increasing (Lehto et al. 2005) and, for instance in 2002, 5760 primary total hip arthroplasties (THA) and 6121 primary total knee arthroplasties were performed in Finland (Mikkola et al. 2005). This corresponds to 112 hip and 118 knee operations per 100 000 inhabitants.

Total hip replacement is an accepted and widely used procedure for advanced osteoarthritis of the hip. It relieves symptoms of pain and helps to restore the loss of function that follows hip arthritis. The most common indications for total hip replacement are osteoarthritis, fracture, rheumatoid arthritis and aseptic bone necrosis (Towheed & Hochberg 1996). It is primarily done to improve HRQoL rather than to extend life (Chang et al. 1996). The favourable effect of hip and knee replacement surgery on HRQoL is well established (Rissanen et al. 1995; Dawson et al. 1996, Norman-Taylor et al. 1996; Rissanen et al. 1996, Towheed & Hochberg 1996). In addition, several studies have suggested that total hip and knee arthroplasty are both relatively cost-effective compared to many other health care interventions (Laupacis et al. 1994; James et al. 1996; Rissanen et al. 1997; Garellick et al. 1998; O'Shea et al. 2002; Segal et al. 2004). The estimates of cost-effectiveness, however, have usually been based on results of individual studies comparing only two or a few interventions at a time. The widely held view of the favourable cost-utility ratio of THA or TKA, compared to many other interventions, is still subject to some uncertainty.

In the Department of Orthopaedics at HUCH, data collection began in March 2002 and ended in August 2002. A total of 358 consecutive patients admitted for scheduled hospital treatment were invited to participate. Approximately six and 12 months after the operation, a follow-up questionnaire was mailed to all patients who had returned the first questionnaire. After data collection had commenced, and after excluding patients whose intervention was not for hip or knee replacement or who had returned only one of the two postoperative questionnaires, 223 patients were available for analysis. In 2002, the total number of hip and knee replacement operations matching our criteria in the HUCH was 1089 (46 patients underwent two operations). Our sample thus represents approximately 33% of the hip and knee replacements performed in the HUCH per year.

## 4.4 Effectiveness of the interventions

The effectiveness of the interventions in terms of change in HRQoL score was measured by the 15D. For spinal surgery patients, the baseline 15D HRQoL questionnaires were mailed together with the invitation to attend the hospital. The cataract and orthopaedic patients were given the questionnaire by the nurses during the pre-operative visit to the hospital. The follow-up questionnaire was mailed to all patients having returned the baseline questionnaire three to 12 months after the operation, depending on the medical specialty studied. The timing of the repeat questionnaire was adjusted so that the full benefits of the operation were expected to have materialised at the time of the reassessment of HRQoL. All recruited patients received treatment according to normal hospital routines.

In addition to HRQoL, outcomes in the orthopaedic patients were measured by two commonly used disease-specific measures, the Harris Hip Score (HHS) (Harris 1969; Söderman & Malchau 2001) and the Knee Society Score (KSS) (Insall et al. 1989; The Knee Society 2006), at baseline and at 12 months after the operation.

In the cataract patient group the HRQoL in terms of the 15D profile and score at baseline and at follow-up was compared to that of the general population standardised according to the age and gender distribution of each separate patient group. The HRQoL data of the general population measured by the 15D came from the Health 2000 Health Examination Survey (Aromaa & Koskinen 2004).

## 4.5 Costs

The perspective taken for the analysis was that of the provider of secondary health care. Direct hospital costs (actual costs of production of the service from the hospital accounting records) relating to the observation period were obtained from the Ecomed® clinical patient administration system (Datawell Ltd., Finland), where all costs of treatment of individual patients in the hospital are routinely stored. Productivity costs for example due to possible periods of disability were not included. Cost-utility was calculated as the relation of QALYs gained to costs of providing surgery. QALYs for CUA were calculated using the obtained utility data and the expected remaining life years of the patients (Statistical Yearbook of Finland 2003).

## 4.6 Statistics

Data were analysed using the statistical software SPSS for Windows, version 11.0 (SPSS Inc., Chicago, IL, USA). In study IV, the R environment for statistical computing and graphics was also used (Ihaka & Gentleman 1996).

The results are given as mean  $\pm$ SD or as mean and 95 % confidence interval (CI) or as median. The statistical significance of differences between pre- and post treatment level values of the 15D dimensions and 15D scores was tested using the Student paired-sample t-test. An independent samples t-test was used to compare the level values and 15D scores with age- and gender-standardised general population. A p-value of  $< 0.05$  was considered statistically significant.

The change in 15D score was defined as the difference between the end-of-study score and the baseline score. In cases where there were three or fewer missing answers on the dimensions of the 15D, the missing responses on any dimension of the 15D were predicted by regression models with the responses on the other dimensions, age and gender, as explanatory variables.

One-way sensitivity analyses were performed by varying the discount rate between 1–5 %, using the median values of QALY gain and costs, and using the upper and lower values of the 95 % CI for the mean differences in treatment effectiveness (HRQoL change) and costs. Cost-utility analyses were based on cost per QALY gained using the 15D as the basis for calculating QALYs.

In study III, to evaluate whether the HRQoL change differed between patients whose surgery was performed within 60 days of the preoperative visit and those whose surgery was delayed until after 60 days, an analysis of variance was performed.

In study IV the relationship between the level value of the 15D dimension of seeing and visual acuity in the eye to be operated on and the non-surgical eye was assessed by Spearman correlation. Visual acuity data obtained by the Snellen chart were converted to Logarithm of Minimal Angle of Resolution (LogMAR) units for statistical analysis using the Visual Acuity Conversion Chart (Visual Acuity Conversion Chart, 2005). To assess the degree of uncertainty 10000 re-samples from the original stochastic cost-utility data set were simulated using a bootstrapping technique. Bootstrap results are presented as cost-effectiveness planes and cost-effectiveness acceptability curves. The cost-effectiveness acceptability curve is a method for summarising the uncertainty in estimates of cost-effectiveness.

In study V the  $\chi^2$ -test was used to compare percentage distributions between the groups.

## 4.7 Ethical Considerations

All patients received the previously scheduled routine treatment and other than being asked to fill in the 15D questionnaire and to give a written informed consent, they were not approached in any other way. To avoid possible unnecessary distress to the relatives of the patients, whether the patients were still alive was ascertained from the Finnish Population Register Centre prior to mailing the repeat questionnaires. The Ethical Committee of the Helsinki and Uusimaa Hospital District approved the study protocol (registration number 26/E6/02). The trial has been registered in the Helsinki and Uusimaa Hospital District Clinical Trials Register ([www.hus.fi](http://www.hus.fi)) with the unique trial number 75370.



## 5 RESULTS

### 5.1 Systematic review of the literature

The literature search identified 4 878 publications (Figure 2). However, 996 were reviews, letters, or editorials and, given that we were looking for original studies, were not included for further review. Furthermore, the 996 articles were excluded if they included publications dealing with prevention or screening, topics which had been decided to be excluded from the review. Thus we were left with 3882 articles altogether that potentially reported QALYs as outcome measures. After screening of abstracts, 624 full-text articles were selected for closer inspection. Of them 72 (representing 70 separate studies) were deemed to fulfil the selection criteria and were included in the review. In 80 cases (13 % of the 624 full-text articles) the initial evaluation by the two independent reviewers differed regarding whether the article was based on clearly identifiable HRQoL data obtained with a valid instrument or not. In those cases the article was also evaluated by a third evaluator and the final decision was made in a consensus meeting. Of those 80 articles, 18 were finally deemed to merit inclusion in the review.



FIGURE 2. Flowchart showing the selection of articles

The most frequently used HRQoL instrument in the identified studies was EQ-5D, which had been used in 47 % of studies, followed by HUI (8.8 %), the Rosser-Kind Index (6.3 %), QWB (6.3 %), SF-6D (5.0 %), and 15D (2.5 %). The rest (24 %) used a direct valuation method (TTO, SG, VAS or RS). The most often studied medical specialties were orthopaedics (15.5 %) followed by pulmonary diseases (13 %), and cardiology (9 %). 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. Ninety percent of the studies came from four countries: the United Kingdom, the United States, Canada and the Netherlands. Approximately half of the articles were randomised trials with high-quality methodologies. An economic analysis was included in 86 % of studies and four of them were based on economic modelling. The mean number of QALYs gained by various interventions varied widely as did the costs per QALY gained. Nearly half of the studies concluded that the studied intervention was cost-effective.

## 5.2 Feasibility

During the study period considered here (from March 2002 to March 2004) 4900 patients returned the first questionnaire and 2680 the follow-up questionnaire. The average response rate to the first questionnaire was over 70 percent, but there was some variation between different clinical specialties. To date (January 2007) more than 11 000 patients with more than 30 different medical entities have completed the first questionnaire.

Routine collection of HRQoL data as an indicator of treatment effectiveness proved to be feasible, requiring only a small amount of extra work and being potentially useful when combined with existing measures of hospital performance. Based on the costs (consisting of salaries, printing, copying, mailing, data recording, statistical consulting, meeting and travel costs, development of necessary software for patient administration) of the first 23 months (approximately €73 000) the cost per patient recruited was approximately €16 in the early phases of the project.

During the study the routine of sending out questionnaires could be automated in such a way that data collection required only a limited amount of extra work. Patients appeared to be generally satisfied with the fact that the hospital was interested in their well-being after treatment. No physician who was offered the chance to participate refused to allow data collection in the patient group he or she was responsible for. The attitudes of the nursing staff were generally positive towards data collection, although it caused some more extra work for some of them.

### 5.3 Neurosurgery

Spinal surgery led to a statistically significant and clinically important improvement in mean HRQoL score in the studied patients (Table 2). At the three-month follow-up the mean HRQoL score (on a 0-1 scale) had increased in a statistically significant manner both in the cervical group, and in the lumbar group (Table 2). Of the 15 dimensions of the instrument, sleeping, usual activities, discomfort and symptoms, depression, distress, vitality and sexual activity had all improved significantly at the three-month follow-up in both groups ( $p < 0.05$ ). In addition, moving and elimination improved statistically significantly ( $p < 0.05$ ) in the lumbar group.

The mean ( $\pm$  SD) waiting time to operation was 68 days ( $\pm$  57, range 0-388) for the cervical spine patients and 73 days ( $\pm$  51, range 2-226) for the lumbar spine patients. In the 88 cervical spine patients that waited less than 60 days for the operation, the HRQoL score increased more (0.05 vs. 0.02) than in the 81 patients that waited more than 60 days. A similar trend was also seen in lumbar spine patients, where the HRQoL score improvement was greater for those 43 patients that waited less than 60 days than for those that waited longer (0.08 vs. 0.05, respectively).

### 5.4 Cataract surgery

The mean HRQoL score had not statistically significantly increased six months after cataract surgery in the whole sample (Table 2). For further analysis, the cataract surgery patients were divided into three subgroups: group A: only one eye was operated on ( $n = 87$ ), group B: both eyes were operated during the follow-up ( $n = 73$ ), and group C: the first eye had been operated on earlier, prior to the study, and now the second eye was operated on ( $n = 59$ ). Of the subgroups, only those who had had both eyes operated on (group B) showed a statistically significant ( $p < 0.001$ ) improvement in the mean HRQoL ( $\pm$  SD) score from 0.80 ( $\pm$  0.13) to 0.83 ( $\pm$  0.14). In group A, the HRQoL remained almost constant and in group C it deteriorated.

Of the 15 dimensions of the 15D instrument, only seeing improved significantly after the operation ( $p < 0.001$ ). 17 % of patients reported having no preoperative difficulties in seeing and 47 % only minor difficulties (levels 1 and 2 of the five-level seeing dimension). Their mean HRQoL score did not improve as a result of surgery. In patients reporting significant or major preoperative seeing problems (levels 3 to 5 of the seeing dimension), cataract surgery improved seeing ( $p < 0.001$ ) and distress ( $p = 0.036$ ), and also had a statistically significant positive effect on the overall HRQoL score, which increased from 0.76 ( $\pm$  0.14) to 0.78 ( $\pm$  0.15) ( $p = 0.02$ ).

The correlation between the best-corrected visual acuity (expressed in LogMAR units) in the surgical eye and the subjective level of seeing (the seeing dimension of the HRQoL instrument) was poor ( $R = 0.17$ ,  $p = 0.013$ ). However, the visual acuity of the non-surgical eye correlated fairly well with the seeing dimension of the 15D instrument ( $R = 0.503$ ,  $p < 0.001$ ).

Compared to age- and gender-standardised general population, the cataract patients were preoperatively statistically significantly worse off on the dimensions of seeing, moving, sleeping, usual activities, depression and distress, but better off on the dimension of mental function. However, the overall HRQoL score did not differ in a statistically significant manner between the general population and the patients.

## 5.5 Hip and knee replacement surgery

The mean HRQoL score increased statistically significantly both in primary hip and in knee replacement patients from the preoperative status to 12 months (Table 2). In revision hip replacement patients, the mean increase was not statistically significant. Of the 15 dimensions of the 15D instrument, moving, usual activities, discomfort and symptoms, distress and vitality had improved statistically significantly in both primary replacement groups. The revision operation improved only the dimension of moving.

Those with a favourable result from primary THA ( $\geq 0.03$  increase in the HRQoL score) were significantly younger than patients with a less optimal result. In primary TKA patients only shorter waiting time was significantly associated with favourable outcome.

The mean ( $\pm$ SD) HHS score of the primary THA patients before operation was 51 ( $\pm 16$ ) and that of the revision patients 47 ( $\pm 21$ ). Twelve months after the operation the scores had increased to 81 ( $\pm 33$ ) ( $p < 0.001$ ) and 74 ( $\pm 35$ ) ( $p = 0.005$ ), respectively. The KSS of the primary TKA patients before operation was 43 ( $\pm 17$ ) and had increased to 88 ( $\pm 27$ ) at 12 months ( $p < 0.001$ ).

## 5.6 Costs of interventions

In cervical spine surgery the mean hospital costs were €3 356 and in lumbar spine surgery €3 493. In the cataract surgery group the mean hospital costs were €1 318 in group A, €2 289 in group B and €1 323 in group C. The mean direct hospital costs of primary THA and TKA were fairly similar, but the mean cost of revision hip replacement operations was clearly higher (Table 2).

TABLE 2. Baseline characteristics and follow-up HRQoL and cost data. (\*\*\*) denotes significant ( $p < 0.001$ ) improvement in HRQoL score from baseline to follow-up). (THA = total hip arthroplasty; Revision THA = revision total hip arthroplasty; TKA = total knee arthroplasty; Cervical = cervical spine surgery; Lumbar = lumbar spine surgery; Cataract = cataract surgery).

	THA (n = 96)	Revision THA (n = 24)	TKA (n = 103)	Cervical (n = 169)	Lumbar (n = 101)	Cataract (n = 219)
Mean age, years (SD)	63 (12)	69 (10)	69 (11)	52 (10)	55 (15)	71 (11)
Female, %	63	50	75	41	59	65
Mean HRQoL at baseline (SD)	0.81 (0.08)	0.81 (0.09)	0.81 (0.09)	0.81 (0.11)	0.79 (0.10)	0.82 (0.13)
Mean HRQoL at follow-up (SD)	0.86 (0.12) ***	0.82 (0.10) N.S.	0.84 (0.11) ***	0.85 (0.11) ***	0.85 (0.12) ***	0.83 (0.14) N.S.
Mean hospital costs, € (SD)	8 737 (2 786)	11 239 (2 996)	8 047 (2 051)	3 356 (603)	3 493 (2 463)	1 643 (530)

## 5.7 Quality adjusted life-years gained by the studied interventions

The mean QALY gain was largest in lumbar spine surgery 2.010 ( $\pm 2.958$ ) followed by THA 1.302 ( $\pm 2.621$ ), cervical spine surgery 1.210 ( $\pm 2.969$ ), TKA 0.575 ( $\pm 1.228$ ), revision THA 0.215 ( $\pm 0.831$ ), and cataract surgery 0.207 ( $\pm 1.002$ ).

The mean discounted (5 %) QALY gain was largest for lumbar spine surgery followed by primary hip arthroplasty, cervical spine surgery, primary knee arthroplasty, revision hip arthroplasty, and cataract surgery (Figure 3).

In cervical spine surgery, the mean QALY gain in patients that waited less than 60 days for the operation was clearly greater than in patients that waited more than 60 days (1.68 vs. 0.70). A similar trend was also seen in lumbar spine patients, where the number of QALYs produced by the intervention (2.50 vs. 1.64) was higher for patients that waited for less than 60 days.

In the cataract patients, the mean QALY gain was clearly highest in those who had had both eyes operated on during the follow-up, 0.446 ( $\pm 1.197$ ). In patients that had had only one eye operated on, it was 0.161 ( $\pm 0.942$ ) and in those whose first eye had been operated on earlier it was -0.022 ( $\pm 0.742$ ).

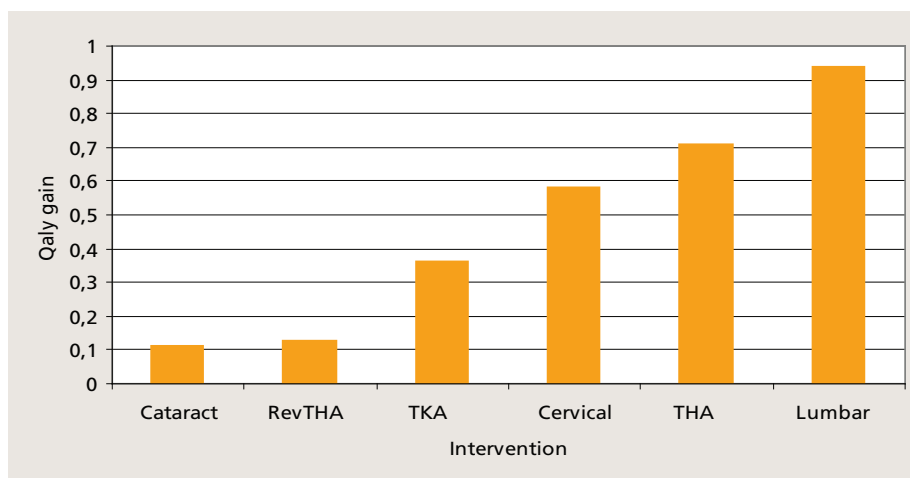


FIGURE 3. Mean number of QALYs (discounted by 5%) gained by various surgical interventions (Cataract = cataract surgery; RevTHA = Revision total hip arthroplasty; TKA = total knee arthroplasty; Cervical = cervical spine surgery; THA = total hip arthroplasty; Lumbar = lumbar spine surgery).

## 5.8 Cost per quality-adjusted life year

The cost per QALY gained was €2 770 for cervical, and €1 740 for lumbar spine operations, respectively. Prolonged waiting for an operation doubled the cost per QALY. For cataract surgery the cost per QALY gained in the whole patient sample was €7 947. In subgroup A the cost per QALY was €8 210 and in subgroup B €5 130, respectively. In subgroup C the cost per QALY gained could not be established as the change in the HRQoL score was negative. In arthroplasty patients the cost per QALY gained was most favourable for primary THA (€6 710), followed by primary TKA (€14 000) and revision THA (€52 270) (Figure 4).

A comparison of all surgical interventions reported here reveals that both the undiscounted and the discounted (5 %) cost per QALY gained was lowest for lumbar spine surgery followed by cervical spine surgery, primary hip arthroplasty, cataract surgery, primary knee arthroplasty and revision hip arthroplasty (Figure 4). However, in patients undergoing cervical or lumbar spine surgery, the cost per QALY was dependent on the waiting time for the operation, with longer waiting times approximately doubling the cost per QALY.

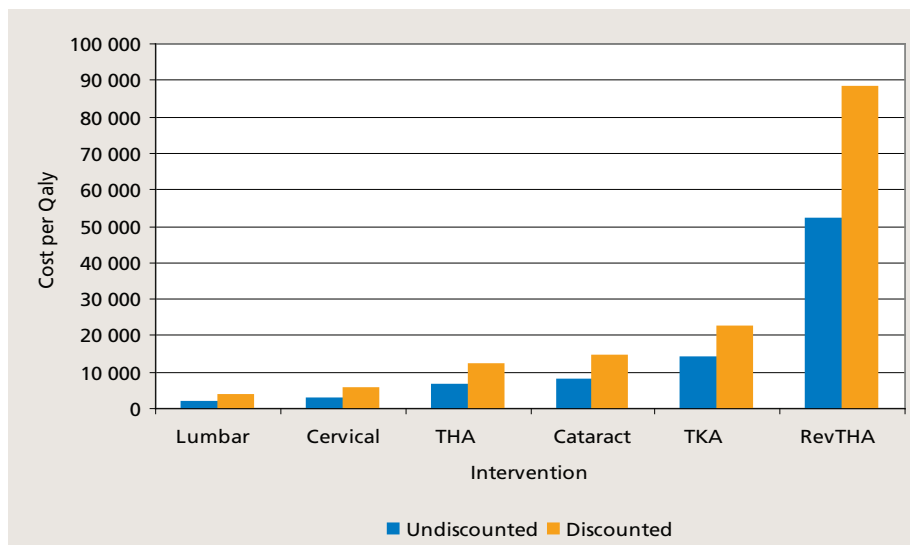


Figure 4. Cost per QALY (undiscounted and discounted by 5 %) gained by five surgical interventions (Lumbar=lumbar spine surgery; Cervical=cervical spine surgery; THA=total hip arthroplasty; Cataract=cataract surgery; TKA=total knee arthroplasty; RevTHA=Revision total hip arthroplasty).

## 5.9 Sensitivity analyses

In cervical and lumbar spine patients (Study III) the cost per QALY gained was relatively robust against discounting and against both varying treatment cost or treatment effectiveness (HRQoL) within the 95 % confidence interval of the means observed in the study.

Similarly, in cataract patients (Study IV) the cost per QALY was in the one-way sensitivity analysis robust against discounting or varying the cost or effectiveness of treatment within the 95 % confidence interval observed in the study. However, using median values increased the cost per QALY substantially in the group of patients whose first eye had been operated on earlier.

In cataract patients, bootstrap simulation suggested that compared to no treatment, surgery was more costly and less effective in 46.4 % of simulated cases, and more costly and more effective in 53.6 % of simulated cases in subgroup A. The corresponding percentages were 37.9 % and 62.1 % in subgroup B and 51.1 % and 48.9 % in subgroup C, respectively. Bootstrap sensitivity analysis also suggested that at a willingness to pay threshold of €20000 per QALY gained, the probability of cataract surgery being acceptable was 51.7 % in subgroup A, 59.0 % in subgroup B and 46.4 % in subgroup C.

In the arthroplasty patients (Study V) the cost per QALY gained by primary THA ranged from €4 770 to €12 340, and that of primary TKA from €9 870 to €24 020 depending on the assumptions used in the analysis. Revision hip replacement would produce a negative QALY gain precluding the calculation of cost per QALY, if effectiveness was at the lower boundary of the 95 % CI of the mean observed in the study. In contrast, if the effectiveness of revision hip replacement were at the higher boundary of the 95 % CI of the mean observed, the cost per QALY would – although still being rather high – go down to €19 860.



## 6 DISCUSSION

Scarce resources are a reality in secondary health care and will, at least in the long run, necessitate prioritisation between different patient groups and interventions. Such prioritisation, if it is to be fair and balanced, is not possible without a systematic, continuous evaluation of the effectiveness and costs of various treatments. The main aim of this study was to test the feasibility of an approach to the routine collection of cost-effectiveness data and to establish the potential benefits such a system could provide. A similar approach has earlier been used in Finland in two pilot studies (Blom-Lange 1998; Kukkonen 2005). Furthermore, numerous studies from other parts of the world have evaluated the cost-utility of the treatment of individual diseases. However, this is to our knowledge the first large-scale experiment aimed at evaluating the cost-effectiveness of several medical interventions side-to-side.

The approach taken in this study turned out to be feasible. The estimation of treatment effectiveness using systematic HRQoL assessments did not require significant extra efforts or investments and the attitudes of the hospital personnel were generally positive towards the project. The latter, for instance, is exemplified by the fact that a great majority (92 %) of the orthopaedic patients were offered the 15D questionnaire despite the fact that this meant some extra work for the nurses responsible for patient care. No physician offered the chance to participate refused to allow data collection in the clinical specialty he/she is responsible for. In all cases, the initiation of data collection naturally required personal communication with those responsible for distribution of the first questionnaires, and in some cases continuous attention and motivation was needed to ensure successful continuation of the data collection.

Patients also appreciated the somewhat unusual event that the hospital showed interest in their well-being, indicated by the fairly high response rate to the first questionnaire compared to surveys in general. Even the response rate for the second questionnaire was good at around 80 %.

To sufficiently automate data collection, even though it remained paper-based throughout the initial phase of the project, some investments in software solutions were necessary at the beginning of the project. To enable mailing of the questionnaires, a Microsoft Access® database was tailored according to the needs of the project, which allowed sufficient automation so that mailing could easily be handled by one person alone. However, to reduce the amount of human resources needed, large scale data collection in the future should, at least in part, be internet-based so that patients can enter the data directly into a database.

The merging of effectiveness data with cost data and information on the treatments was unproblematic in the Ecomed® database. However, in cases where individual patients had been treated multiple times during the follow-up, it was

sometimes difficult to ascertain to which treatment the effectiveness data were connected. Similarly, it is also virtually impossible to tell the effect of a single intervention on the change in HRQoL if the patient has received treatments for more than one illness during the follow-up. Some of the problems can be overcome by manually checking that the effectiveness data are linked to the right treatment during the follow-up period, but this requires extra work and is feasible only when the study sample size is small. It does not, however, solve the problem of establishing how much each treatment has contributed to the change in HRQoL. In large patient samples, hand checking may not be that important as some deviant patients easily disappear into the mass if thousands of patients are studied and thus do not bias the results in a significant way.

The primary strength of the current economic analysis is that it is based on real world patients of the HUCH receiving scheduled surgical treatment under currently accepted practices and indications. Other published economic analyses have mostly been based on clinical trial data, which may not reflect the situation observed under routine circumstances. Our results, therefore, reflect the true effectiveness of interventions instead of merely their efficacy. Admittedly, randomised controlled trials estimating the efficacy of various new treatments are also essential when considering whether to introduce a new treatment or not. However, results from an RCT represent at best the efficacy of an intervention, not its effectiveness (Gold et al. 1996). Subsequent effectiveness data are thus essential if decisions on the allocation of resources are to be as objective and comprehensive as possible. Consequently, there will always be a need for both types of methodologies – that of RCT and real-life studies – to be carried out sequentially and if necessary, repeatedly over time.

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 health care resources is based on scientific evidence on the value of various interventions regarding their ability to produce societal welfare.

A strength of the QALY approach used in this study is the fact it reflects patients' own assessment of their health status. The key issue in assessing the effectiveness of secondary health care should be the patient's own evaluation: how they feel that the intervention has affected their HRQoL. Traditionally, the providers of treatments have assessed the effectiveness of interventions by using provider-oriented outcomes. Such an approach is, however, prone to lead to subjective and biased assessments. Even though health care 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 adequately. Furthermore, the clinical outcomes used to evaluate effectiveness are usually disease- or at best specialty-specific, and

do not allow a comparison of treatment results across different clinical specialties. Therefore, studies based on real patient data are arguably of much more value for the decision-maker pondering over allocation of resources.

For instance, regarding cataract surgery, the evaluation of visual acuity or visual quality of life alone, although important, serves only the purpose of a health status measure, and is not sufficient to reflect the true HRQoL of patients. According to some recent studies the impact of seeing on generic HRQoL may have been exaggerated in the past. Although some studies have established a relationship between difficulties in seeing and HRQoL, others have indicated that visual impairment does not affect generic HRQoL as much as generally assumed. For instance in patients with unilateral visual impairment, a deteriorated HRQoL – as measured by the SF-36 instrument – was found only in those moderately to severely affected (Chia et al. 2004). Espallargues et al. (2005) have reported that for patients with age-related macular disease, even severe visual impairment was not reflected in patients' own assessment of life and satisfaction measured using several HRQoL instruments. Furthermore, the disease-specific VF-14 score was found to correlate only moderately with visual acuity in the better eye (Steinberg et al. 1994).

The cost per QALY gained for knee arthroplasty was two-fold compared to hip replacement. For revision hip arthroplasty the cost per QALY was clearly higher than for the studied primary interventions. The latter finding stresses the importance of the primary operation and the choice of the best possible implant material so that revision operations can be avoided as far as possible.

This study showed that there is great variation in the cost-utility of various interventions performed in a real-world setting even when only common, widely accepted interventions are considered. Spinal surgery turned out to be the most cost-effective intervention regarding the cost per QALY gained. On the other hand, spinal disorders are usually self-limiting and most patients recover with time also without surgery. To establish the real incremental cost-effectiveness of spinal surgery comparative studies employing conservative treatment as a control would be necessary.

Lansingh et al. (2007) reported in a recent meta-analysis that the cost-utility of cataract surgery varies greatly worldwide. In their opinion, that is mainly a consequence of differing methods and time frames used for assessing the benefits of the intervention. According to their results, cataract surgery is comparable to hip arthroplasty in respect of cost-effectiveness, and more cost-effective than knee arthroplasty or implantation of a defibrillator. In our material, cataract surgery was more expensive regarding the cost per QALY than seen in many earlier studies. One explanation for this might be the mixed patient material of a university clinic with many patients suffering from secondary ophthalmic problems in addition to cataract. Another reason for the unexpectedly small increase in HRQoL after cataract surgery may be that two thirds of the patients reported only minimal

preoperative subjective difficulties in seeing despite objective evidence of poor visual acuity in the operated eye. The small HRQoL gain observed as the result of routine cataract surgery was confined mainly to an improvement in seeing only. In patients suffering from significant or major preoperative difficulties in seeing, the HRQoL gain was more encouraging. It is possible that cataract operations have earlier been performed at a later stage of the disease and on patients that have suffered more from the detrimental consequences, such as injuries (Stenevi et al. 2000) due to poor visual acuity. Consequently, the gain achieved by the operation might also have been larger in earlier studies. From our current results it is clear, however, that to justify resource use on cataract surgery, the patient has to have definitive medical indications for the surgery or its cost-effectiveness needs to be proven with clear paybacks in the form of improved quality of life.

In addition to providing the means for comparing the cost-effectiveness of interventions performed in different medical specialties, our approach also makes it possible for surgeons and surgical units within one specialty to compare their performance with those of their colleagues, and thus hopefully improve their practices and generate clinical benefits for the patients. When generalizing these results to other hospital districts, one has to be cautious because of possible variation in costs of treatment.

Further research is urgently required to give more comparators for analysis and better evidence as a background for decisions. This calls for continuous monitoring of the cost-utility of various interventions to optimise health care spending. Changes to the delivery and organisation of health services should be evaluated before they are widely implemented. Too much momentum may lead to an inappropriate implementation of change before an evaluation is complete (McDonnell et al. 2006).

Cost-effectiveness analysis is not the only circumstance where HRQoL assessments are useful. They can also be used to refine patient selection, to evaluate which patient groups gain most from interventions, and to compare different service providers in terms of effectiveness of care.

## 6.1 Limitations of the study

Cost data used for the analyses covered only the relevant specialty-related costs and not the costs of primary health care and outpatient medication or productivity costs. Ideally the perspective for an economic analysis should be societal and take into account all types of costs, not only direct hospital costs. However, Brouwer et al. (2006) have suggested that it may be useful to adopt a two-perspective approach as a standard, presenting one cost-effectiveness ratio following a strict health-care perspective and one following the common societal perspective.

In at least two of the studied interventions, the costs of hospital treatment make up the bulk of the direct total costs of treatment, as there is no known effective cure for cataract or osteoarthritis of the hip or knee other than surgery. In the case of back surgery this not as evident and it is true that a more detailed analysis of the costs of conservative treatment would be useful, but was not within the scope of the current study. Thus the method of using cost data readily available from the hospital accounting system, instead of tediously collecting data from individual patients, can be defended, and is in a trial that involves several thousand patients the only feasible possibility as yet to acquire reasonably reliable cost data. In the future, however, emphasis needs to be put on efforts to develop methods to reliably estimate the costs of the whole treatment chain (including at least the costs of primary care and those of medication) by for instance combining existing registers already containing the necessary information. A nationwide standardised electronic system for routine recording of resource use and HRQoL data at all levels of the health care system would be of immense importance for more comprehensive cost data collection and routine assessment of cost-effectiveness of treatment.

It is possible that the use of a societal perspective could affect some of the conclusions of this study and certainly would have some impact on the cost per QALY figures. The extent of the impact is, however, impossible to establish presently. Already ongoing studies may answer some of these questions in the future. Preliminary results from such studies employing arthroplasty patients similar to those enrolled in our study indicate that the use of additional health care services, such as home care or rehabilitation, is in arthroplasty patients fairly limited and thus do not have a major effect on the overall health care costs of an arthroplasty patient (Hirvonen et al. 2006).

The health economic analytical approach used in this study can be criticised for making some fairly straightforward assumptions. One of them is the assumption that the studied interventions do not affect the length of life. Also the lack of multivariate analysis to handle the confounding factors that may be present in routine-styled settings can be criticised. On the other hand, the relevance of some economic evaluations has been criticised for the lack of real world data and poor generalisability of results (Guidelines for the Economic Evaluation of Health Technologies, 2006), which may especially be a problem in health economic modelling often based on vaguely defined data. Nonetheless, this is probably the first endeavour to evaluate the cost-utility of a large part of secondary health care and, although the methods of data collection and analysis certainly need to be refined in the future, it provides a sound, pragmatic basis for the start of continuous, routine monitoring of the cost-utility of secondary health care.

The time horizon of data collection ranged in our study from three to 12 months. Regarding the interventions studied here, this can be considered sufficient as all of the studied interventions can be expected to have produced maximal benefit by the time of repeat data collection. For some other health care

interventions this may not always be true and the timing of data collection needs to be adjusted individually to meet the needs of specific research questions in other fields of medicine.

Our QALY calculations are based on the simplistic assumption that the improvement in HRQoL is permanent. This is of course not always true as HRQoL is known to deteriorate with advancing age while other concurrent diseases may also affect it negatively. In the future, longer follow-up times would be needed to make the analyses more reliable. In some cases more refined analyses using health economic modelling using registry data—for example the number of re-interventions in orthopaedics – may certainly be necessary. The HRQoL data collected in our experiment provides a sound basis for such analyses. On the other hand, health economic modelling is always based on at least some assumptions regarding the natural course of the studied disease and is thus also subject to uncertainties. Our results, by contrast, reflect the true everyday effectiveness accomplished in health care. Although our findings must be interpreted with care, they represent, in our opinion, as valid an approach to defining the cost-utility of secondary care as health economic modelling.

## 6.2 Future perspectives

Although the results of the project provide a sound basis for comparing the cost-utility of various interventions, another question is whether they really influence health care decision-making. There are indications from earlier studies that feedback of good quality effectiveness data is not always able to influence for instance the decisions of practicing surgeons (Wright et al. 2002). A future objective should thus be to study the use of the cost-utility data produced in the project, and to establish whether such data really are able to influence medical decision making in a significant manner.

Another future priority should be to study whether the observed poor outcome in some patients can be predicted by pre-treatment HRQoL dimensions or available clinical variables. These analyses can partly be performed using already collected data; there is a need to also use patient registers. In the future, methods based on artificial intelligence to predict treatment outcomes on a case-by-case basis may replace current approaches based mainly on mean data obtained from groups of patients. Such data could help tailor treatment approaches according to the needs and prognoses of individual patients (Ryynänen 2005).

## 7 CONCLUSIONS

On the basis of the present study, the following conclusions can be drawn.

1. Although QALYs 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 with a valid method remains fairly limited.
2. Routine collection of HRQoL data as an indicator of treatment effectiveness is feasible, requires only a small amount of extra work, and is potentially useful when combined with existing measures of hospital performance.
3. Spinal surgery led to a statistically significant and clinically important improvement in HRQoL and the cost per QALY gained was reasonable; however a prolonged delay in the surgical intervention led to an approximate doubling of the cost per QALY gained by the treatment.
4. Cataract surgery had only a minor effect on HRQoL in most of the patients and the improvement was confined mostly to patients whose both eyes were operated on during the study period. The cost of cataract surgery per QALY was much higher than previously reported.
5. Primary hip and knee replacement were both effective in improving the mean HRQoL score in our patients. By contrast, the improvement in HRQoL after revision hip arthroplasty was fairly small and statistically insignificant. The cost per QALY gained from knee replacement is twice that of hip replacement.

## Acknowledgements

The present study was carried out at the Hospital District of Helsinki and Uusimaa, during the years 2002–2006.

I want to thank everyone who helped and supported me during this project. My warm thanks go to the staff and patients of the Hospital District of Helsinki and Uusimaa. Without their participation this study could not have succeeded.

In particular, I want to express my deepest thanks to the following persons:

The supervisors of this thesis, Professor Harri Sintonen, Ph.D. and Docent Risto P. Roine, M.D., Ph.D. whose major contribution to this work has been invaluable. I respect the fact that they shared their outstanding knowledge and experience in the field of health economics and their scientific and medical knowledge with me. Their interest in my thesis was very encouraging. They have always had time to teach me. I have been remarkably privileged to have had the possibility to know them and work together with them.

The reviewers of this thesis, Professor Martti Kekomäki, M.D., Ph.D. and Research Professor Markku Pekurinen, Ph.D., for their valuable comments and constructive criticism that helped me to improve the text.

The other members of the HUS QoL Study Group: Docent Marja Blom, Ph.D., Anne Heikkilä, M.Sc., Professor Olli-Pekka Rynänen, M.D., Ph.D., Virpi Semberg-Konttinen, M.Sc. and Bo-Gustav Wilson, M.Sc. (Economics) for their encouragement and interest in this work which has been of great value. The invaluable contribution of Professor Pekka Laippala, Ph.D. (who died on May 13, 2003) during the early phases of the project is also gratefully acknowledged.

The other co-authors of the original publications of this thesis, docent Kari Krootila, M.D., Ph.D., Tiina Leivo, M.D., Ph.D., Professor Pekka Paavolainen, M.D., Ph.D., Eija Roine, BM, and Professor Juha Öhman, M.D., Ph.D., for their efforts in writing the articles and good collaboration.

I owe my thanks to Anna-Maija Koivisto, M.Sc. for sharing her expertise in statistics that helped in finding the correct methods for analysing the data. I am grateful to Merja Jauhiainen, M.Sc. for her decisive contributions to the literature searching.

Development manager Mikko Rotonen, Senior Planning Officer Petri Mikkola, Product Manager Jussi Järveläinen, Information Specialist Jouni Vierumäki, Senior Systems Analyst Tiina Alikoski for help in questions regarding data handling and storage. Mrs Marja Nurmi for technical assistance. My sincere thanks to Mark Phillips, M.Sc. for editing the language of this thesis.

My present boss Professor Marjukka Mäkelä, M.D., Ph.D., and all the other wonderful colleagues and co-workers at Finohta for their support during this work.



All my relatives and friends for keeping me in touch with other perspectives of life.

And finally, I want to thank Jari my loving husband, who has supported me in all possible ways and our children Lotta and Lauri for giving me so much love and happiness.

This study was financially supported by grants from the Hospital District of Helsinki and Uusimaa, the Yrjö Jahnsson Foundation and the Finnish Office for Health Technology Assessment.

Helsinki, May 2, 2007

Pirjo Räsänen

## References

- Angevine P, Zivin J, McCormick, P. Cost-effectiveness of single-level anterior cervical discectomy and fusion for cervical spondylosis. *Spine* 2005;30:1989–97.
- Apajasalo M, Sintonen H, Holmberg C, Sinkkonen J, Aalberg V, Pihko H, Siimes MA, Kaitila I, Mäkelä A, Rantakari K, Anttila R, Rautonen J (a). Quality of life in early adolescence: a sixteendimensional health-related measure (16D). *Qual Life Res* 1996;5:205–11.
- Apajasalo M, Rautonen J, Holmberg C, Sinkkonen J, Aalberg V, Pihko H, Siimes MA, Kaitila I, Erkkilä K, Sintonen H (b). Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res* 1996;5:532–8.
- Aromaa A, Koskinen S. Health and functional capacity in Finland. Baseline results of the Health 2000 health examination survey. Publications of National Public Health Institute 2004;Series B 12/2004:171.
- Arnesen T, Nord E. The value of DALY life: problems with ethics and validity of disability adjusted life years. *BMJ* 1999;319:1423–5.
- Bertalanffy H, Eggert HR. Clinical long-term results of anterior discectomy without fusion for treatment of cervical radiculopathy and myelopathy. A follow-up of 164 cases. *Acta Neurochir* 1988;90:127–35.
- Blom-Lange Marja. Elämänlaadun mittaamismahdollisuuksista erikoissairaanhoidossa – väli-raportti tutkimuksen toteutettavuudesta [On the possibilities of measurement of health-related quality of life in secondary health care – interim report on the feasibility of the study]. Päijät-Hämeen sairaanhoitopiirin julkaisuja A 3/1998. Lahti 1998.
- Blom-Lange M, Sintonen H. Measuring the effectiveness of hospital treatment. A paper presented at the 19th Nordic Health Economists' Study Group Meeting in Oslo, Norway, August 21–22, 1998. p. 1–10. SINTEF Unimed. 1998.
- Blom-Lange M, Sintonen H. The effectiveness of hospital treatment in terms of health-related quality of life. A paper presented at the 21th Nordic Health Economists' Study Group Meeting in Lund, Sweden, August 25–26, 2000. <http://www.nek.lu.se/ryde/luche99/Papers/blom.pdf>.
- Bozic KJ, Rosenberg AG, Huckman RS, Herndon JH. Economic evaluation in orthopedics. *The Journal of Bone and Joint Surgery* 2003;8:129–42.
- Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873–84.
- Brouwer WB, van Exel NJ, Baltussen RM, Rutten FF. A dollar is a dollar is a dollar – or is it? *Value Health* 2006;9:341–7.
- Bucciero A, Vizioli L, Cerillo A. Soft cervical disc herniation. An analysis of 187 cases. *J Neurosurg Sci* 1998;42:125–30.
- Chang RW, Pellissier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *JAMA* 1996;275:858–65.
- Chia EM, Wang JJ, Rochtchina E, Smith W, Cumming RR, Mitchell P. Impact of bilateral visual impairment on health-related quality of life: the Blue Mountains Eye Study. *Invest Ophthalmol Vis Sci* 2004;45:71–6.
- Cost-effectiveness analysis registry. <http://www.tufts-nemc.org/cearegistry/> [Referred May 2005]
- Cox HL, Laupland KB, Manns BJ. Economic evaluation in critical care medicine. *Journal of Critical Care* 2006;21:117–24.
- Daffner SD, Hilibrand AS, Hanscom BS, Brislin BT, Vaccaro AR, Albert TJ. Impact of neck and arm pain on overall health status. *Spine* 2003;28:2030–5.
- Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care* 1996;5:81–8.
- Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 2004;13:437–52.
- Dowie J. Analysing health outcomes. *J Med Ethics* 2001;27:245–50.
- Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996;313:275–83.

- Drummond MF, O'Brien BJ, Stoddard GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Second Edition. Oxford University Press, 1997.
- Drummond MF, McGuire A (eds). *Economic Evaluation in Health Care – Merging Theory with Practice*. Oxford University Press, 2001.
- Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL, *Methods for the Economic Evaluation of Health Care Programmes*. Third Edition. Oxford University Press, 2005.
- Espallargues M, Czoski-Murray CJ, Bansback NJ, Carlton J, Lewis GM, Hughes LA, Brand CS, Brazier JE. The impact of age-related macular degeneration on health status utility values. *Invest Ophthalmol Vis Sci* 2005;46:4016–23.
- Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:1–74.
- Gaetani P, Tancioni F, Spanu G, Rodriguez y Baena R. Anterior cervical discectomy: an analysis on clinical long-term results in 153 cases. *J Neurosurg Sci* 1995;39:211–8.
- Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 1997;16:1–31.
- Garellick G, Malchau H, Herberts P, Hansson E, Axelsson H, Hansson. Life expectancy and cost utility after total hip replacement. *Clin Orthop* 1998;346:141–51.
- Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy* 1992;21:249–79.
- Gerard K, Smoker I, Seymour J. Raising the quality of cost-utility analyses: lessons learnt and still to learn. *Health Policy* 1999;46:217–38.
- Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. Oxford University Press, 1996.
- Goldberg EJ, Singh K, Van U, Garretson R, An HS. Comparing outcomes of anterior cervical discectomy and fusion in workman's versus non-workman's compensation population. *Spine* 2002;27:408–14.
- Greenberg D, Pliskin JS. Preference-based outcome measures in cost-utility analyses. *Int J Technol Assess Health Care* 2002;18:461–66.
- Guide to the Methods of Technology Appraisal. National Institute for Clinical Excellence 2004.
- Guidelines for the Economic Evaluation of Health Technologies: Canada, 3rd Edition, 2006. [http://www.cadth.ca/media/pdf/186\\_EconomicGuidelines\\_e.pdf](http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf) [Referred December 2006]
- Guidelines for economic evaluation of pharmaceuticals, 2nd edition: Canada. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA) 1997.
- Hailey D, Ohinmaa A, Roine R (a). Evidence for the benefits of telecardiology applications: a systematic review. Alberta Heritage Foundation for Medical Research, 2004.
- Hailey D, Ohinmaa A, Roine R (b). Study quality and evidence of benefit in recent assessments of telemedicine. *J Telemed Telecare* 2004;10:318–24.
- 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;51:737–55.
- Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med* 2001;33:358–70.
- Hay J. Health care costs and outcomes: how should we evaluate real world data? *Value in Health* 1999;2:417–9.
- Hirvonen J, Blom M, Tuominen U, Seitsalo S, Paavolainen P, Lehto MUK, Hietaniemi K, Rissanen P, Sintonen H. Lonkan ja polven tekonivelleikkaukseen jonottavien potilaiden sekä verrokkiväestön terveys- ja sosiaalipalvelujen käytön vertailua [Comparison of the use of health care and social services between patients waiting for hip and knee replacement surgery and the general population]. *Suomen Ortopedia ja Traumatologia* 2006;3:350–2.
- Hospital District of Helsinki and Uusimaa. <http://www.hus.fi/> [Referred January 2007]
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996, 5:299–314.
- Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of The Knee Society clinical rating system. *Clin Orthop* 1989;248:13–4.
- James M, St Leger S, Roswell KV. Prioritising elective care: a cost utility analysis of orthopaedics in the north west of England. *J Epidemiol Community Health* 1996;50:182–9.

- Järvelin J. Health Care Systems in Transition - Finland. European Observatory on Health Care Systems. 2002;4:1 <http://www.euro.who.int/document/e74071.pdf> [Referred December 2006]
- Kauppinen R. Cost-effectiveness of intensive patient education for guided self-management of adult novel asthmatic patients. Doctoral dissertation. Kuopio: University Printing Office, 1999.
- Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-based medicine: How to Review and Apply Findings of Healthcare Research. The Royal Society of Medicine Press, 2003.
- Kristensen FB, Horder M, Poulsen PB (eds.). Health technology assessment handbook. 1st ed. Copenhagen: Danish Centre for Evaluation and Health Technology Assessment; 2001. <http://www.sst.dk/Applikationer/cemtv/publikationer/docs/Metodehaandbog/MethodologyHandbook180601.pdf> [Referred December 2006]
- Kukkonen J. Terveystieteiden tutkimuslaitoksen tutkimusraportteja. [An evaluation of the effectiveness of public health services on the basis of routinely collected data]. Doctoral dissertation, Kuopio University Publications D. Medical Sciences 355, 2005.
- Käypä hoito -suositus [Current Care Guideline]. Aikuisiän harmaakaihi [Cataract in adult-age] [www.kaypahoito.fi](http://www.kaypahoito.fi).
- Lansingh VC, Carter MJ, Martens M. Global cost-effectiveness of cataract surgery. *Ophthalmology*. 2007 Mar 23; [Epub ahead of print]
- Larsen RJ, Asmussen M, Christensen T, Olsen J. Economic evaluations in international health technology assessment - A study of methodologies. *Danish Health Technology Assessment* 2003;5(1).
- Laupacis A, Bourne R, Robareck C, Feeny F, Wong C, Tugwell P, Leslie K, Bullas R. Costs of elective total hip arthroplasty during the first year. Cemented versus noncemented. *J Arthroplasty* 1994;9:481-7.
- Lehto MUK, Jämsen E, Rissanen P. Lonkan ja polven endoproteesikirurgia – varaosien avulla liikkujaksi. [Endoprosthesis surgery of hip and knee, spare parts improve mobility]. *Duodecim* 2005;121:893-901.
- McDonnell A, Wilson R, Goodacre S. Evaluating and implementing new services. *BMJ* 2006;332:109-12.
- Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for base utility. *Health Econ* 2005;14:487-96.
- Mikkola H, Järvelin J, Seitsalo S, Keskimäki I. Ortopediset leikkaukset Suomessa 1987-2002. [Orthopedic surgeries in Finland 1987-2002. Number of surgeries according to the geographic areas, waiting time and centralization]. *Duodecim* 2005;121:861-71.
- Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med* 2000;132:964-72.
- Nord E. Cost-Value Analysis in Health Care. Making Sense out of QALYs. New York: Cambridge University Press, 1999.
- Norman-Taylor FH, Palmer CR, Villar RN. Quality-of-life improvement compared after hip and knee replacement. *J Bone Joint Surg Br* 1996;78:74-77.
- O'Shea K, Bale E, Murray P. Cost analysis of primary total hip replacement. *Ir Med J* 2002;95:177-80.
- Palmer S, Torgerson DJ. Definitions of efficiency. *BMJ* 1999;318:1136.
- Petitti DB. Meta-analysis, Decision Analysis and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. Oxford University Press, 2000.
- Punnonen H. Sairaaloitosten ja sairaanhoitopiirien tammi-elokuu 2006 [January to August 2006 in hospitals and hospital districts]. Suomen Kuntaliitto, 2006.
- Raftery J. Economics evaluation: an introduction. *BMJ* 1998;316:1013-4.
- Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ* 2004;329:224-7.
- Rawlins M, Dillon A. NICE discrimination. *J Med Ethics* 2005;31:683-4.
- Rissanen P, Aro S, Slätis P, Sintonen H, Paavolainen P. Health and quality of life before and after hip or knee arthroplasty. *J Arthroplasty* 1995;10:169-75.

- Rissanen P, Aro S, Sintonen H, Slätis P, Paavolainen P. Quality of life and functional ability in hip and knee replacements: a prospective study. *Qual Life Res* 1996;5:56–64.
- Rissanen P, Aro S, Sintonen H, Asikainen K, Slätis P, Paavolainen P. Costs and cost-effectiveness in hip and knee replacements. A prospective study. *Int J Technol Assess Health Care* 1997;13:575–88.
- Ryynänen OP. Pisteillä hoitoon. [Prioritizing instead of waiting in line] *Duodecim* 2005;121:349–50.
- Sampath P, Bendebba M, Davis JD, Ducker T. Outcome of patients treated for cervical myelopathy. A prospective, multicenter study with independent clinical review. *Spine* 2000;25:670–6.
- Segal L, Day SE, Chapman AB, Osborne RH. Can we reduce disease burden from osteoarthritis? *Med J Aust* 2004;180:S11–17.
- Shiell A, Donaldson C, Mitton C, Currie G. Health economic evaluation. *J Epidemiol Community Health* 2002;56:85–8.
- Sintonen H. An approach to measuring and valuing health states. *Soc Sci Med* 1981;15 C: 55–65.
- Sintonen H, Pekurinen M. A generic 15 dimensional measure of health-related quality of life (15D). *J Soc Med* 1989;26:85–96.
- Sintonen H, Pekurinen M. A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In *Quality of Life Assessment: Key Issues in the 1990s*. Walker SR, Rosser RM (eds). Kluwer Academic Publishers, Dordrecht 1993,185–195.
- Sintonen H. Outcome measurement in acid-related diseases. *PharmacoEconomic* 1994;5:17–26.
- Sintonen H: The 15-D measure of health-related quality of life: II Feasibility, reliability and validity of its valuation system. Working Paper 42. National Center for Health Program Evaluation, Melbourne, 1995. <http://www.buseco.monash.edu.au/centres/che/pubs/wp42.pdf> [Referred December 2006]
- Sintonen H. Kustannusvaikuttavuus erikoissairaanhoidossa [Cost-effectiveness in specialized health care]. In Mikkola T., Outinen M. (eds.) *Kohti ensi vuosituhat. Näkymiä erikoissairaanhoidon laadunhallintaan*. Helsinki Stakes Raportteja 1996;205:32–37.
- Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med* 2001;33:328–336.
- Sintonen H: Taloudellinen arviointi (Economic evaluation). In: *Menetelmien arviointi terveydenhuollossa*. [Assessment of technologies in health care]. Mäkelä M, Kaila M, Lampe K, Teikari M (eds.) Kustannus Oy Duodecim, Helsinki, 2007, pp 90–115.
- Statistical Yearbook of Finland 2003. Statistics Finland, Helsinki, 2003.
- Stavem K. Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Qual Life Res* 1999;8:45–54.
- Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM. National study of cataract surgery outcomes. Variation in 4-month postoperative outcomes as reflected in multiple outcome measures. *Ophthalmology* 1994;101:1131–41.
- Stenevi U, Lundström M, Thorburn W. The cost of cataract patients awaiting surgery. *Acta Ophthalmol Scand* 2000;78:703–5.
- Stirling B, Roberts T, Heginbotham C, McCallum A. QALY-maximisation and public preferences: results from a general population survey. *Health Econ* 2002;11:679–693.
- Stone PW, Chapman RH, Sandberg EA, Liljas B, Neumann PJ. Measuring costs in cost-utility analyses. Variations in the literature. *Int J Technol Assess Health Care* 2000;16:111–24.
- Söderman P, Malchau H. Is the Harris Hip Score system useful to study the outcome of total hip replacement. *Clin Orthop Rel Res* 2001;384:189–197.
- The 15D© health-related quality of life (HRQoL) instrument home page. <http://www.15d-instrument.net/15D> [Referred January 2007]
- The Knee Society. <http://www.kneesociety.org/index.asp/fuseaction/site.outcomes> [Referred May 2006]
- Tosteson AN, Gabriel SE, Kneeland TS, Moncur MM, Manganiello PD, Schiff I, Ettinger B, Melton LJ III. Has the impact of hormone replacement therapy on health-related quality of life been undervalued? *J Womens Health Gend Based Med* 2000;9:119–130.

- Towheed TE, Hochberg MC. Health-related quality of life after total hip replacement. *Semin Arthritis Rheum* 1996;26:483–91.
- Tugwell P, de Savigny D, Hawker G, Robinson V. Applying clinical epidemiological methods to health equity: the equity effectiveness loop. *BMJ* 2006;332:358–61.
- Visual Acuity Conversion Chart. *J Cataract Refract Surg* 2005;31:A4.
- Ware J, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF36 health profile and summary measures: summary of results from the Medical Outcome study. *Med Care* 1995;33:264–79.
- Wright CJ, Chambers GK, Robens-Paradise Y. Evaluation of indications for and outcomes of elective surgery. *CMAJ* 2002;167:461–6.
- Zentner A, Valasco-Garrido M, Busse R. Methods for the comparative evaluation of pharmaceuticals. *GMS Health Technol Assess*, 1:Doc09, 2005. <http://www.egms.de/pdf/journals/hta/2005-1/hta000009.pdf> [Referred January 2007]