Hjelt Institute, Department of Public Health,
and
Hospital District of Helsinki and Uusimaa, Group administration
Faculty of Medicine, University of Helsinki

Measuring health-related quality of life (HRQoL) and quality-adjusted life years (QALY) in the critical care setting

Tarja Vainiola

ACADEMIC DISSERTATION
To be presented, with the permission of the Faculty of Medicine of the University of Helsinki,
for public examination in Auditorium XV, University Main Building,
Unioninkatu 34, on May 24th 2014

Helsinki 2014

Supervised by:

Professor (emeritus) Harri Sintonen, Ph.D.
Hjelt Institute
Department of Public Health
University of Helsinki

Professor Risto P. Roine, M.D., Ph.D.
University of Eastern Finland
Department of Health and Social Management
Research Centre for Comparative Effectiveness and Patient Safety (RECEPS)

Chief Physician
Helsinki and Uusimaa Hospital District
Research and Development

Reviewed by:

Professor Tero Ala-Kokko, M.D., Ph.D.
Oulu University Hospital
Department of Anesthesiology
Division of Intensive Care Medicine
Oulu, Finland

Adjunct Professor Juha Laine, Ph.D.
Pfizer Oy
Helsinki, Finland

Official opponent:

Professor (emeritus) Martti Kekomäki, Ph.D.
Faculty of Medicine
University of Helsinki

ISBN 978-952-10-9810-9 (Paperback)
ISBN 978-952-10-9811-6 (PDF)

Unigrafia, Helsinki, 2014

To Aada, Lukas and grandchildren yet to come

Table of Contents

Summary

Background: Cost-utility analysis provides a means to determine the health benefit and economic burden of different health-care interventions. In cost-utility analyses, the benefit of care is measured in quality-adjusted life years (QALYs) gained. The calculation of QALYs requires knowledge of the change in health-related quality of life (HRQoL) and assumptions concerning when the benefit of care materialises and how long the benefit lasts. The gold standard for QALY calculations has not yet been defined and, as a consequence, the HRQoL instruments and calculation methods used vary from study to study.

Aims: The aim of the current study was to clarify how much the differences in the components used for the calculation of QALYs are reflected in the end result, i.e., the number of QALYs gained in the critical care setting. The detailed aims were to study 1) the effect of the instrument used (the EQ-5D or the 15D) on the HRQoL score and the measured changes in it; 2) the effects of the baseline HRQoL and the assumptions concerning the progress of recovery on the number of QALYs; 3) how to estimate life expectancy in the critical care setting, and 4) which factors have an effect on the follow-up HRQoL.

Patients and methods: The results are based on two study populations. The first population comprises patients having been treated in an intensive care or high-dependency unit (N = 3600), and whose HRQoL was assessed using the EQ-5D and 15D HRQoL instruments 6 and 12 months after treatment. The second population consists of patients having undergone treatment in a cardiac surgery intensive care unit (N = 980), and whose HRQoL was assessed using the 15D HRQoL instrument at baseline, when placed on a waiting list for surgery and 6 months after treatment.

Main results: The results of the studies show that the HRQoL index score is dependent on the instrument used. The distribution of the patients' HRQoL scores differed between instruments. The differences are explained, inter alia, by the ceiling effect of the EQ-5D—i.e., for a significant proportion of the respondents, the instrument produced the best possible HRQoL score of 1—and by the negative scores of the EQ-5D — i.e., for health states worse than death. The 15D produced higher mean HRQoL scores than the EQ-5D. The 15D was able to distinguish between a greater number of health states than the EQ-5D, thus showing a better discriminatory power.

The choice of instrument was also reflected in the change observed in HRQoL. The two instruments classified patients according to the change in HRQoL (improved, remained stable, deteriorated) in a similar manner only in approximately half of the cases. The 15D was more sensitive to detecting a change than the EQ-5D. Consequently, both its discriminatory power and responsiveness to change were better than those for the EQ-5D.

The assumptions concerning the progression of recovery and the baseline HRQoL score had an effect on the number of QALYs gained both within and between instruments and, consequently, on the cost per QALY ratio. The EQ-5D and the 15D performed differently under different calculation assumptions. The greatest difference in the number of QALYs gained was caused by the negative HRQoL scores observed with the EQ-5D enabling the accrual of more than 1 QALY per year.

Patients having been treated in an intensive care unit showed long-lasting excess mortality and, as a consequence, a reduced life expectancy. By contrast, in cardiac surgery patients, the life expectancy was similar to or even better than that of the general population. In patient groups with excess mortality, neither the follow-up time nor the life expectancy of the general population can be regarded as optimal indicators for the duration of the benefit of care. In those patient groups, life expectancy should be extrapolated in relation to the observed excess mortality.

In cardiac surgery patients, factors predicting mortality and morbidity are not able to accurately predict the follow-up HRQoL. Instead, patient experiences, such as restlessness and pain during intensive care, predicted poor post-treatment HRQoL. Given that these results are novel, future studies should be directed to patient experiences during treatment. They may be confounding factors in analyses concerning treatment effectiveness, and also diminish the effectiveness of treatment.

Conclusion: QALY is not a universal measure, but is dependent on the HRQoL instrument used and on how the factors to be taken into account in the calculation of QALYs are chosen and defined. Furthermore, factors external to the interventions under evaluation, such as the patient's psychological experiences during treatment, may have an effect on the follow-up HRQoL. The ranking of different interventions in terms of their effectiveness calls for standardisation in the calculation of QALYs and more information on the effect of patient experiences during treatment on the follow-up HRQoL

Tiivistelmä

Tausta: Kustannus-utiliteettianalyysien avulla voidaan selvittää terveydenhuollon eri interventioiden terveyshyötyjä ja kustannuksia. Näissä analyyseissa hoidon hyödyt mitataan laatupainotettuina lisäelinvuosina (QALY). QALY:n laskemiseen tarvitaan tieto terveyteen liittyvän elämänlaadun muutoksesta sekä oletus toipumisen kulusta ja hoidon tuottaman hyödyn kestosta. Toistaiseksi ei ole määritelty kultaista standardia QALY:n laskemiselle, minkä seurauksena käytetyt elämänlaatumittarit ja laskentatavat vaihtelevat.

Tavoite: Tämän tutkimuksen tarkoituksena on selvittää, missä määrin erot laatupainotetun elinvuoden laskemisen osatekijöissä vaikuttavat saavutettujen laatupainotettujen elinvuosien määrään tehohoitoympäristössä. Tutkimuksen yksityiskohtaiset tavoitteet ovat selvittää 1) tuottavatko EQ-5D ja 15D samanlaisen arvion potilaiden terveyteen liittyvästä elämänlaadusta ja elämänlaadun muutoksesta; 2) mikä on lähtötilanteen elämänlaadun ja toipumisen kulkuun liittyvien erilaisten oletusten vaikutus saavutettujen laatupainotettujen elinvuosien määrään; 3) miten elinajanodote tulisi arvioida tehohoitopotilailla; ja 4) mitkä tekijät ennustavat seuranta-ajan terveyteen liittyvään elämänlaatua.

Aineisto ja menetelmät: Väitöskirja koostuu kahdesta aineistosta. Ensimmäinen aineisto käsittää teho- ja valvontaosastoilla hoidettuja potilaita (n=3600), joiden terveyteen liittyvä elämänlaatu mitattiin EQ-5D ja 15D elämänlaatumittareilla kuusi ja 12 kuukautta hoidon jälkeen. Toinen aineisto koostuu sydänkirurgian teho-osastolla hoidetuista potilaista (n=980), joiden terveyteen liittyvä elämänlaatu mitattiin 15D elämänlaatumittarilla hoitojonoon asettamisen yhteydessä ja kuusi kuukautta hoidon jälkeen.

Päätulokset: Tutkimuksen tuloksena todettiin, että terveyteen liittyvää elämänlaatua osoittava indeksiluku on riippuvainen käytetystä elämänlaatumittarista. Mittarien tuottamien elämänlaatuindeksien jakaumat erosivat toisistaan. EQ-5D:llä oli taipumus kattoefektiin eli varsin suuri osa vastaajista sai mittarilla maksimiarvon (=1), mikä tarkoittaa, että terveyteen liittyvä elämänlaatu olisi paras mahdollinen (täysin terve). Lisäksi EQ-5D tuotti negatiivisia elämänlaatuindeksejä eli elämänlaadun tiloja, jotka kuvastavat kuolemaa heikompaa elämänlaatua. 15D tuotti keskimääräisesti korkeampia elämänlaadun indeksejä. Se pystyi paremmin erottelemaan elämänlaadun eri tiloja kuin EQ-5D eli sen erottelukyky oli parempi.

Käytetty elämänlaatumittari vaikutti myös terveyteen liittyvässä elämänlaadussa havaittuun muutokseen. Mittarit luokittelivat potilaat elämänlaadun muutoksen suhteen (parantunut, ennallaan, heikentynyt) samalla lailla noin puolessa tapauksista. 15D oli herkempi havaitsemaan elämänlaadun muutosta kuin EQ-5D. Näin ollen sekä erottelukyky että muutosvaste olivat 15D:llä parempia kuin EQ-5D:llä.

Oletukset toipumisen kulusta ja käytetty lähtötilanteen elämänlaatuindeksin arvo vaikuttivat mittarien sisällä ja välillä saavutettuihin laatupainotettuihin elinvuosiin ja sitä kautta kustannukset/QALY -suhteeseen. EQ-5D ja 15D toimivat erilailla eri laskentaoletuksilla. Suurimman eron saavutettuihin laatupainotettuihin elinvuosiin aiheutti EQ-5D:n negatiiviset elämänlaatuindeksit arvot, jotka mahdollistavat enemmän kuin yhden laatupainotetun elinvuoden kertymisen vuoden aikana.

Tehohoitopotilailla havaittiin kauan jatkuva ylikuolleisuus väestöön verrattuna ja ylikuolleisuuden seurauksena alentunut elinajanodote. Sydänkirurgisilla potilailla elinajanodote sen sijaan vastasi väestön elinajanodotetta, tai oli jopa sitä parempi. Tautiryhmissä, joissa havaitaan ylikuolleisuutta, seuranta-aika ja väestön elinajanodote eivät ole optimaalisia hoidon hyödyn keston suureita. Näissä tautiryhmissä elinajanodote tulisi ekstrapoloida suhteessa havaittuun ylikuolleisuuteen.

Sydänkirurgisilla potilailla sairastavuutta ja kuolleisuutta kuvastavat indikaattorit eivät ennustaneet seuranta-ajan terveyteen liittyvää elämänlaatua. Sen sijaan tehohoidon aikainen levottomuus ja kivuliaisuus ennustivat

hoidon jälkeistä alentunutta elämänlaatua. Koska tulokset ovat uusia, tulee tutkimusta suunnata edellä mainittuihin ja muihin potilaan hoidon aikaisiin kokemuksiin, jotka saattavat olla sekoittavia tekijöitä hoidon vaikuttavuutta arvioitaessa ja heikentää hoidon tuloksellisuutta.

Johtopäätös: Laatupainotetut elinvuodet eivät ole universaali mittayksikkö, vaan riippuvainen käytetystä elämänlaatumittarista ja siitä, miten laskennassa huomioon otettavat osatekijät on määritelty ja valittu. Lisäksi arvioitavien hoitomuotojen ulkopuoliset tekijät, kuten potilaan hoidonaikaiset kokemukset, saattavat vaikuttaa koettuun elämänlaadun muutoksen. Eri hoitomuotojen asettaminen paremmuusjärjestykseen vaikuttavuuden suhteen edellyttää laatupainotettujen elinvuosien laskennan standardointia ja lisää tietoa potilaiden hoidonaikaisten kokemusten vaikutuksesta elämänlaadun muutokseen.

List of original publications

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I Vainiola T, Pettilä V, Roine RP, Räsänen P, Rissanen AM, Sintonen H. Comparison of two utility instruments, the EQ-5D and the 15D, in the critical care setting. Intensive Care Med 2010;36:2090-3.

II Vainiola T, Roine RP, Pettilä V, Kantola T, Räsänen P, Sintonen H. Effect of health-related quality-of-life instrument and quality-adjusted life year calculation method on the number of quality-adjusted life years gained in the critical care setting. Value Health 2011;14:1130-4.

III Vainiola T, Roine RP, Suojaranta-Ylinen R, Vento A, Sintonen H. Can factors related to mortality be used to predict the follow-up health-related quality of life (HRQoL) in cardiac surgery patients? Intensive Crit Care Nurs 2013;29:337-43.

IV Vainiola T, Seppä K, Roine RP, Notkola I-L, Suojaranta-Ylinen R, Sintonen H. The estimated life expectancy of critical care patients: The effect of excess mortality and length of follow-up. (Submitted)

Abbreviations

AQoL	Assessment of Quality of Life
AUC	Area under the curve
AVR	Aortic valve replacement
BMI	Body mass index
CABG	Coronary Artery Bypass Graft Surgery
CABG+Valve	Combined Coronary Artery Bypass Graft and valve Surgery
CSICU	Cardiac surgical intensive care unit
CUA	Cost-utility analysis
DCE	Discrete Choice Experiment
EQ-5D	EuroQol
EQ-5D-3L	EQ-5D 3 level
EQ-5D-5L	EQ-5D 5 level
ES	Effect size
EuroSCORE	European method for Cardiac Operative Risk Evaluation
HDU	High dependency units
HRQoL	Health-related quality of life
HUI	Health Utilities Index
ICD-10	International Classification of Diseases, 10^{th} edition
ICU	Intensive care unit
LR-test	Likelihood ratio test
MCID	Minimal clinically important difference
MG	Magnitude Estimation
MID	Minimal important difference
MM-OC	Median TTO valuations
MVR	Mitral valve replacement
NICE	National Institute for Health and Care Excellence (United Kingdom)
NYHA	New York Heart Association Functional classification
OPCAB	Off-pump Coronary Artery Bypass
PTO	Person Trade-Off
QALY	Quality-adjusted life year
RASS	Richmond Agitation and Sedation Scale
RRT	Renal replacement therapy
RS	Rating Scale
RSR	Relative survival ratio
SD	Standard deviation
SEM	Standard error of mean
SG	Standard Gamble
SOFA	Sequential Organ Failure Assessment score
SRM	Standardized response mean
TISS-28	Therapeutic Intervention Scoring System
TTO	Time Trade-Off
VAS	Visual Analogue Scale
WHO	World Health Organization
VRS	Verbal Rating Scale
WTD	Worse than death
WTP	Willingness-to pay

1. Introduction

The development of new treatment methods and the reorganisation of functions have led to an ongoing change in the organisation of health-care services. Many interventions formerly requiring inpatient care are now performed as day surgeries while increasingly demanding interventions become available. As a consequence, a continually growing number of hospital days comprise of days spent in an intensive care environment. For example, in the United States between 2000–2005, the number of acute care hospital beds decreased and, at the same time, the number of critical care beds increased, resulting in a situation where 15% of all hospital beds were situated in a critical care environment. As a consequence, the number of critical care inpatient days increased by 10.6% (Halpern and Pastores, 2010).

In addition to the reorganisation of services, the ageing of the population also increases the demand for critical care. The elderly (≥65 years of age) use more critical care compared to the younger population (<65 years). In Minnesota, the elderly were reported to have used a mean of 125.3 ICU days/1 000 person years compared to 17.1 ICU days/1 000 person years in the younger population (Seferian and Afessa, 2006). In Finland, the annual growth of hospital days in critical care setting has been concentrated on the elderly population (older than 65 years). This trend has been steadily increasing in recent years (Figure 1). The elderly account for about 40% of all inpatient days in critical care settings (Intensium benchmarking database).

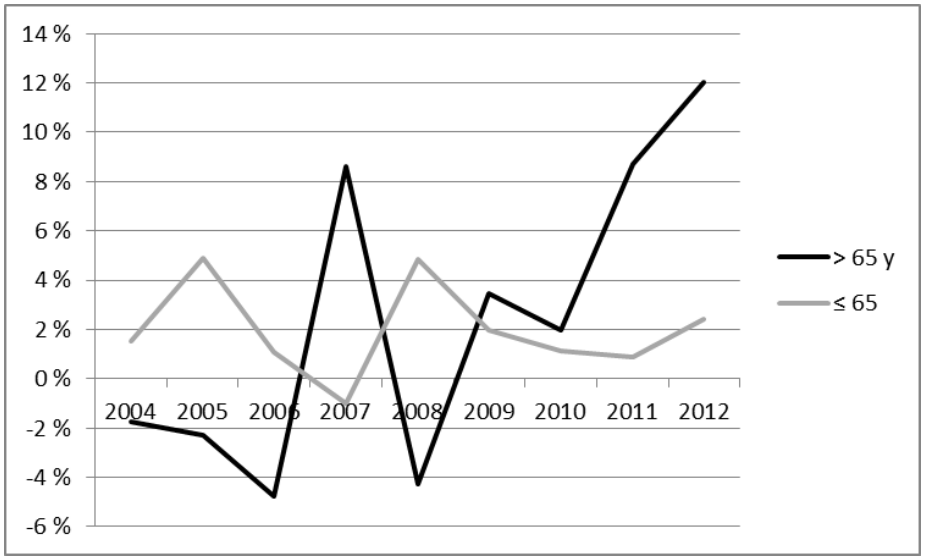

Figure 1. Growth rate of hospital days in the critical care setting by age in Finland, 2004–2012

For example, in Finland, the proportion of individuals 65-years-old and older has grown 21% from 2000 to 2010. In 2040, the number of elderly is predicted to grow by more than 660 000 individuals, i.e., 70% more than the number of elderly in 2010 (Statistics Finland, database). Age as such does not generate a need for intensive care; but, since it is associated with an increased prevalence of chronic illnesses, the ageing of a population leads to an increased need for ICU days (Seferian and Afessa, 2006).

Intensive care requires substantial personnel and financial resources. The cost of an inpatient day in a critical care setting is many fold compared to an inpatient day in a normal ward. For example, the cost of an inpatient day in a normal ward of the Helsinki University Hospital in 2013 ranged from 300 € to 900 €, while in ICU, the range was from 2 700 € to 4 300 €. According to the Intensium benchmarking database, there were about 57 000 inpatient days in critical care settings in 2012. In annual costs, this amounts to about 201 € million (an inpatient day costs 3 500 €) (Intensium benchmarking database).

The increasing demand for intensive care together with its resource intensity mandates the assessment of its health gains measured by QALYs. The calculation of QALYs requires knowledge on patients' health-related quality of life (HRQoL), the change in it, knowledge on or at least assumptions about when the benefit of care materialises and how long the benefit lasts.

HRQoL can be measured by disease-specific or generic HRQoL instruments. But, for QALY calculations and cost-utility analyses, generic HRQoL instruments are recommended since they allow, at least in theory, comparisons between different illnesses and their treatments. Widely known generic HRQoL instruments that produce a single index score on a 0–1 scale required for QALY include the Assessment of Quality of Life (AQoL) (Hawthorne et al., 1999), the Health Utilities Index (HUI1, HUI2 and HUI 3) (Furlong et al., 2001), the SF-6D (Brazier et al., 2002), the EQ-5D (Brooks and the EuroQol Group, 1996) and the 15D (Sintonen, 1994; Sintonen, 1995). As mentioned earlier, in addition to the HRQoL score, the duration of the benefit of care and the assumption about the progress of recovery, i.e., the way the change occurs in the HRQoL score over time, are crucial elements in QALY calculations.

The special characteristics of the measurement of HRQoL and the calculation of QALYs within critical care settings require contemplation. First, although no HRQoL instrument can claim to be the gold standard, the 2002 Brussels Roundtable Consensus Meeting recommended the SF-36 and the EuroQol (EQ-5D) as the preferred HRQoL instruments in the critical care setting (Angus and Carlet, 2003). Second, the baseline HRQoL usually has an effect on the follow-up HRQoL score; but, its measurement or estimation in acutely ill critical care patients is challenging (Manca et al., 2005). Third, the effect of a serious illness on life expectancy is often vague and difficult to establish, which poses a major problem as the time horizon used in QALY calculations, e.g., the remaining lifetime of the patient, is uncertain. The aim of the current study was to clarify how much the differences in the components used for the calculation of QALYs are reflected in the end result, i.e., the number of QALYs gained in the critical care setting.

2. Review of the literature

2.1 Generic, single-index health-related quality of life instruments

Generic, single-index HRQoL instruments have been developed in many countries. The AQoL was developed in Australia at the end of the 1990s (Hawthorne et al., 1999), the EQ-5D by European collaboration in the early 1990s (Brooks and the EuroQol Group, 1996) and the 15D in Finland in the early 1980s (Sintonen, 1994; Sintonen, 1995). The Health Utilities Index (HUI) system, which was developed in Canada in the 1980s (Furlong et al., 2001), comprises three different instruments (HUI1, HUI2 and HUI3), of which HUI2 and HUI3 are complementary and can be used in parallel. The SF-6D is derived from the profile instrument SF-36, which was developed in the United States (Hay and Morales, 2001). The development process of the SF-36 began in the 1980s and the final version was introduced in 1990 (Ware, 2000). The revision and algorithm development for a generic, single-index score HRQoL instrument from the SF-36 was completed in the United Kingdom in the early 2000s (Brazier et al., 2002)

All generic HRQoL instruments consist of two elements: the health state descriptive system and the valuation system of health states defined by the health state descriptive system. In the QALY context, HRQoL must be expressed as a single-index score, where 1 represents full health and 0 represents death; some instruments, however, also produce negative scores, which imply health states worse than death (WTD).

The health state descriptive system

As the expression suggests, the aim of the health state descriptive system is to describe all essential dimensions of health from the viewpoint of HRQoL. In practice, the health state descriptive system is a standardised, self-administered questionnaire. There is no generally accepted theory of HRQoL to determine which dimensions to include in the health state descriptive system**.** Many systems have their roots in the classic definition of the World Health Organization (WHO), according to which "health is a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity" (WHO, 1958). Since this definition is quite broad, the aspects of physical, mental and social well-being have been operationalised in different ways and, as a consequence, the health state descriptive systems differ between the instruments (Sintonen, 1994; Sintonen, 1995; Brooks and EuroQol Group, 1996; Hawthorne et al., 1999; Furlong et al., 2001; Hay and Morales, 2001; Brazier et al., 2002). However, the dimensions and their content should be restricted to those upon which health care can have an effect.

The health state valuation system

The aim of the health state valuation system is to establish population preferences—i.e., quality weights—for the different health states defined by the health state descriptive system. The valuation can take place by following a direct and holistic or indirect approach. Typically, in the direct and holistic approach, the health states to be valued are described in written form in their entirety to those from whom the valuations are elicited (the respondents) who must imagine themselves in those hypothetical states even if the valuation takes place in different ways. Using the indirect approach, the valuation is divided into parts or stages and the final HRQoL scores for different health states are aggregated from the results of those stages.

The most frequently used valuation methods are the Visual Analogue Scale (VAS) also called the Rating Scale (RS), Magnitude Estimation (MG), Standard Gamble (SG), Time Trade-Off (TTO), Person Trade-Off (PTO) and Willingness-to pay (WTP) (Grabbe et al., 1997; Green et al., 2000). Normally, in direct valuation methods, a limited number of relatively simple health states are valued directly and the single-index score for most of the health states is extrapolated by statistical methods (Kopec and Willison, 2003).

The minimal clinically important change

The minimal clinically important change i.e., the magnitude of a change in the HRQoL score that a patient can perceive as a change for better or worse can be regarded as an indicator of the effectiveness of care (Walters and Brazier, 2005). Estimates for it have been derived using two different methods — a distribution-based method or an anchor-based method. Distribution-based measures are effect size (ES), standardised response mean (SRM) and standard error of mean (SEM). When estimating the change using distribution-based methods, the distribution of the data has an effect on the results (King, 2011). In the anchor-based method, the change is combined with an external anchor. The external anchor can be a patient's opinion about the development of HRQoL or a clinical measurement or indicator, which expresses the development of the illness (Wyrwich, 2004). The development of HRQoL is established by asking the patient whether s/he has experienced an improvement, no change or deterioration in her/his health compared with the former measurement point (Browne et al., 2010). The anchor-based method is considered more appropriate for estimating a change than distribution-based methods (Revicki et al., 2008).

In the literature, the minimal clinically important change is usually referred to with the acronym MCID (minimal clinically important difference) or with MID (minimal important difference) (Wyrwich, 2004). Conceptually, however, minimal clinically important change and minimal clinically important difference are distinct. The former refers to a change over time—e.g., in a group of patients—whereas the latter refers to a difference in a cross-section — e.g., between two groups. The former can be estimated using the methods described above, whereas there is no direct method to estimate the latter. Therefore, the latter is considered equal to the former and both are referred to with acronyms MCID or MID (Table 1).

Table 1. Properties of the generic, single index HRQoL instruments

	AQoL[1]	**EQ-5D[2]**	**15D[3]**	**HUI2[4]**	**HUI3[4]**	**SF-6D[5]**
Origin	Australia	Europe	Finland	Canada	Canada	USA
Items	35	5	15	7	8	36[6]
Response levels	4-6	3	5	3-5	5-6	2-6
Range	-0.04-1	-0.59-1	0-1	-0.03-1.00	-0.36-1	0.203-1
Valuing system	TTO	TTO	RS	SG	SG	SG
Different health states	$2.37*10^{23}$	243	$3.1*10^{10}$	24 000	972 000	$8.7*10^{20}$
MID	Not estimated	0.08	0.03	0.05	0.05	0.03

[1]Hawthorne et al, 1999; [2] Walters and Brazier, 2005, [3] Sintonen, 1994, Sintonen 1995; [4]Horman et al, 2003; [5] Ware, 2000; [6]The questionnaire used is SF-36

The HRQoL instruments shown in Table 1 were developed across diverse periods. The tendency has been such that the use of an instrument has been most common in the country in which it was developed (Richardson et al., 2012). Currently, the EQ-5D appears to be the most widely used instrument worldwide (Räsänen, 2006). In addition, as mentioned earlier, the 2002 Brussels Roundtable Consensus Meeting recommended the EQ-5D (and the SF-36) as the preferred HRQoL instruments in the critical care setting. Therefore, it seemed appropriate to more closely examine the EQ-5D and to compare it in the critical care setting to the 15D, which is the most frequently used utility instrument in Finland.

2.1.1 The EQ-5D

The EQ-5D was developed by an international research group, named the EuroQol Group. The EuroQol Group was established in 1987 and included members from Finland, Netherlands, Norway, Sweden and the United Kingdom (Rabin and de Charro, 2001). The original goal of the EuroQol Group was to develop a very simple health state descriptive system, to generate from it a small number of different health states to be valued in a standardised way in a representative population sample in different countries and to determine whether the valuations across countries are similar to one another. The instrument was not intended to be used as a stand-alone measure, but to complement other forms of quality-of-life measurement tools and to facilitate the collection of a common dataset for reference purposes (Brooks, 1996).

The health state descriptive system

The health state descriptive system of the EQ-5D was developed through a conceptual process on the basis of the available HRQoL instruments. Altogether, seven different HRQoL instruments were reviewed during the development process. Of these instruments, the Quality of Well-Being Scale, the Rosser Index and the 15D represented both generic and profile instruments (Coast, 1992; Coons et al., 2000; Sintonen, 2001), while the Sickness Impact Profile, the Nottingham Health Profile and the Health Measurement Questionnaire were simply profile instruments (Cole et al., 1994; Coons et al., 2000). The members of the EuroQol Group presupposed that the forthcoming instrument should include dimensions related to mobility, daily activities, self-care, psychological functioning, social and role performance and pain or other health problems (EQ-5D concepts and methods, 2005).

The prerequisites for development were that the chosen dimensions should be wide in content and suitable for different health states, and that the instrument should be usable by the general population in different health states. The first version consisted of six dimensions (6D) with two to three levels on each dimension. The six dimensions were mobility, self-care, main activity, social relationship, pain and mood. The levels of the dimensions were on an ordinal scale except for the dimension of self-care, which was on a nominal scale. The levels on the dimension of self-care were no problems in self-care, unable to dress independently and unable to eat independently. On the basis of experiences and experiments, a new version was ratified in 1991. It consists of five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The dimension social relationship was excluded. The revision focused on levels as well; all dimensions were changed to ordinal and included three levels: no problems, some or moderate problems and unable or extreme problems. The EQ-5D instrument can generate 243 different health states. The number of different health states was limited in order to enable the use of a holistic valuation method (The EuroQol Group, 1990; Brooks et al., 1996).

The basic idea behind the EuroQol Group was to develop a very simple health state descriptive system, to derive from it a small, standardised set of different health states, to value them in a standardised, holistic way in a representative population sample in different countries and to see whether the valuations are similar or different across countries. The EQ-5D was never intended to be a stand-alone instrument; but, rather, was meant to complement other HRQoL measures (EuroQol Group, 1990; Brooks et al., 1996; Sintonen et al. 2003).

The valuation methods

Primarily, two different valuation methods have been used to generate a single-index score for the health states defined by the EQ-5D descriptive system. These methods are the Visual Analogue Scale (VAS) and the Time Trade-Off method (TTO) (Greiner et al., 2003; Rabin and de Charro, 2001). The VAS method consists of a line with two clearly defined end points (Green et al., 2000). In the EQ-5D VAS valuation system, the end points are the best imaginable (100) and the worst imaginable (0) health state (Brooks et al., 1996; Greiner et al., 2003). In the VAS valuation process, respondents draw a line from boxes describing different, earlier defined health states on the scale ("thermometer") to indicate how good or bad they are. In the TTO method, the respondents choose between two alternatives: x years in full health or a previously defined number of years (e.g., 10) in the health state being valued i.e., how much of the lifetime the respondent is willing to sacrifice in order to achieve a higher quality of life (Dolan, 1997; Green et al., 2000). Comparing these two valuation methods, the TTO produced higher utility weights in mild and moderate health states and considerably lower utility weights in severe health states (Brazier et. al., 1999). In QALY calculations and cost-utility analyses, it has been recommended that the EQ-5D scores defined by the TTO method should be used (Rabin and de Charro, 2001).

Using the TTO method, 43 of the 243 possible health states were valued by 3 337 respondents. Each respondent valued 11 different health states varying from very mild to severe health states. In addition to perfect health (11111), worst possible health state (33333), immediate death and unconsciousness (not defined) were valued. The respondents represented the general population of the United Kingdom. The respondents completed the health states valuation differently for health states considered better and worse than death. In the former case, the respondent chose between 10 years in the health state being valued to be equivalent to a length of time (x) in perfect health. In the latter case, the respondent chose between dying immediately and a length of time (x) in the health state being valued following 10 – x years in perfect health (Dolan et al., 1996).

These 43 health states were used to create a regression model to interpolate an index score for the rest ofhealth states. In the regression model, the constant for any dysfunctional state is -0.081, i.e., when a level of some or moderate problems occurs on any of the dimensions. In addition to the constant, the final score consists of the reduced value of the level from each dimension. As a consequence, no health state can obtain a value between 0.888 and 0.999. In addition, the regression model includes a dummy variable (N3) which means that, if any of the dimensions is at level three, -0.269 is subtracted from the score (Dolan, 1997).

These reductions will cause substantial changes in the single-index scores when one moves from one level to another. Pain has the most significant effect on the utility score—if pain is at level 3 and all other dimensions are at level 1, the single-index score is 0.264. The final score is produced through addition. The scale of the index score is -0.594 – 1, where 1 indicates full health and 0 represents death. Altogether, 84 health states—i.e., 35% of all possible health states—are WTD (Walters and Brazier, 2005). The above applies to the UK TTO tariff. The EQ-5D tariff varies between countries — e.g., in the United States, the lowest utility score in the original D1 tariff is -0.102, while in Spain, the lowest is -0.654 (Heijink et al., 2011). All of these tariffs have been based on mean TTO valuations. However, in the United States, a new tariff based on median TTO valuations (MM–OC model) is now recommended for use. In this tariff, the scores range from -0.81 to 1 (Shaw et al., 2010).

The minimal clinically important difference

The EuroQol Group has not estimated MID for the EQ-5D, so there is no commonly accepted value for MID. The MID estimation was initiated by researchers in the 2000s. Using previously published studies, Walters and Brazier investigated MID derived from both the anchor- and the distribution-based methods for 11 different patient groups. Depending on the patient group, the anchor-based MID varied from -0.011 to 0.139 and the distribution-based MID ranged from 0.11 to 0.17. For the entire patient population, the mean MID was 0.074 and the median MID 0.081 when estimated using the anchor-based method (Walters and Brazier, 2005). In addition, MID has been estimated at 0.08 using distribution- and anchor-based methods in cancer patients (Pickard et al., 2007), from 0.08 to 0.10 using the anchor-based method in multiple myeloma patients (Kvam et al., 2011) and at 0.05 using the distribution-based method in rheumatoid arthritis patients (Marra et al., 2005).

In 2011, the EuroQol Group introduced a new version of the EQ-5D instrument. The differences between the old and new versions lie in the number of levels for each dimension as well as some changes in the wording of previous levels. The number of levels has been increased from three to five, which includes no problems, slight problems, moderate problems, severe problems and extreme problems. The name of the EQ-5D instrument has also been further refined, whereby the three-level version is called the EQ-5D-3L and the five-level version is known as the EQ-5D-5L. It is possible to use a link between the EQ-5D-5L and EQ-5D-3L descriptive systems to create an index score for the EQ-5D-5L health states (EuroQol Group, 2011). Until the new valuation system for the EQ-5D-5L is completed, which combines discrete-choice experiment (DCE) and TTO methods, the instrument should be regarded as a profile instrument since the resulting scores yield the EQ-5D-3L scores. Because of the changes in the health state descriptive system (new levels and partly new wording from the previous version) and the new valuations method, the EQ-5D-5L is in fact a new instrument and comparability with results obtained using the EQ-5D-3L are most likely lost.

2.1.2 The 15D

The development of what is now known as the 15D started in the late 1970s. The original target was to develop a comprehensive HRQoL instrument which could be used both as a profile and generic instrument (Sintonen 1994; Sintonen, 2001).

The health state descriptive system

The first version of the 15D was called the 12D and included 12 dimensions with 4–5 levels on each dimension. The conceptual basis for the health state descriptive system relied on the definition of health by WHO (WHO, 1958). In addition, the 12D was based on dimensions of health considered important in contemporary Finnish health policy documents.

According to feedback from the medical profession, the instrument was too concentrated on physical well-being. Thus, a new version was launched in 1986 and included additional dimensions concerning mental health including depression, distress and pain. This revised instrument was named the 15D.1. The suitability of the 15D.1 regarding its ability to reflect HRQoL was tested among nearly 3 000 individuals. The intention was to determine whether the instrument had too many attributes or if something essential was missing. After revisions based on these initial results and statistical analyses, an updated version (the 15D.2) was launched in 1992. In the 15D.2, the ability to work and social participation were combined into one dimension, labeled “usual activities” and a new dimension on sexual activity was added. In addition, all dimensions were changed to five-level scales in order to increase sensitivity. What is today

referred to as the 15D is actually the 15D.2 and the dimensions are mobility, vision, hearing, breathing, sleeping, eating, speech, excretion, usual activities, mental function, discomfort and symptoms, depression, distress, vitality and sexual activity. The instrument can generate 30.5 billion different health states (Sintonen, 1994; Sintonen, 2001).

The valuation system

Due to the large number of different health states, it is not possible to use direct and holistic valuation methods (Honkalampi and Sintonen, 2010). The health states values are produced indirectly by applying multi-attribute utility theory. The valuation process comprised three stages and was performed on a representative sample of the Finnish adult population. During the first stage, relative importance weights were elicited from the top levels of the 15 dimensions. At the second stage, importance weights were elicited from the lowest levels (5) of the dimensions. The valuation procedure was completed using a 0–100 ratio scale (VAS scale), where 100 was given to the most important dimension, and 0, was assigned if a dimension was not considered important at all. The ratio scale nature of the valuation task was emphasized by placing nine arrows to the right-hand side of the 0–100 scale with a text explaining how the number pointing to an arrow should be interpreted over the range of the scale. For example, an arrow pointing to 90 reads, "9/10 as important as the most important attribute (90% as important as the most important attribute)." The importance weights for the intermediate levels were extrapolated linearly from the weights of the extreme ends in relation to the distance between level values, which were elicited for each dimension during the third stage. In addition to the five levels, the states "unconscious" and "dead" were valued on for every dimension. The preference weight for each level was calculated by multiplying the level weight by the importance weight for the dimension. The most important dimensions for good HRQoL are mental function (i.e., to be able to think clearly and logically), to be able to breathe normally and, to be able to perform usual activities (such as work, leisure and hobbies) normally.

The total score over all dimensions through the 3-stage additive valuation procedure is obtained as follows:

$$v_H = \Sigma I_j(x_j)[w_j(x_j)],$$

where $I_j(x_j)$ is the average relative importance people attach to various levels of dimension j (j = 1, 2,..,15) and $w_j(x_j)$ is the average value people place on various levels of dimension j.

The scale for the single-index score is 0–1, where 1 indicates full health, 0.0162 represents unconsciousness and 0 represents death (Sintonen, 1995).

In a recent study, the 15D scores were compared to the TTO valuation of one's own health among 863 patients representing various levels of severity in different disease groups. At the aggregate level, the 15D and TTO scores had good agreement, although in some patient groups the agreement was not that good (Honkalampi and Sintonen, 2010).

The minimal clinically important difference

MID has been estimated by the developer of the instrument using an anchor-based method among 1 231 patients. The anchor was the patient's experience concerning his or her health state compared to 6 months earlier. MID was reported to be 0.03 (Sintonen, 1994). Recently, MID was re-evaluated and 0.03 was considered a suitable clinically important magnitude of change in the case the magnitude should be equal to both directions. In relation to a positive change (i.e., improvement to HRQoL), it is possible that an even smaller change in the range of 0.02 could be experienced by patients as important. On the other hand,

0.03 may be too small a magnitude of change when HRQoL deteriorates when a negative value for MID should reach 0.05 (Alanne, 2011).

2.1.3 Comparison of the EQ-5D and the 15D in different patient populations

The comparison of results using different utility instruments began as early as the end of the 1990s, intensifying during the 2000s. The instruments can be compared in terms of several properties, where sensitivity may be one of the most important properties.

The sensitivity of an instrument entails two aspects. The first aspect concerns its ability to distinguish between individuals and groups in different health states cross-sectionally (discriminatory power). Second, instruments are evaluated based on their ability to detect changes in individuals or groups over time (responsiveness to a change in one's health status). In addition, different criteria can be used for evaluating an instrument's discriminatory power. First, this refers to the ability of the instrument to detect health problems which can be described by a number of different health states. Second, the discriminatory power refers to the ability of the instrument to detect changes in health. This can be described by the ceiling and floor effects. Furthermore, the properties of the distribution of the scores—e.g., skewness and peakedness—can tell researchers something about the discriminatory power (Sintonen, 1994).

The "ceiling" and "floor" effect and skewness can also be used to describe the instrument's responsiveness to change. In addition, responsiveness indices such as ES and SRM have been used. ES is defined as the change in the mean score from the baseline to follow-up divided by the standard deviation at the baseline measurement. SMR is the mean response divided by the standard deviation of responses, which equals the paired t-statistic without factoring in the sample size (Liang et al., 1990).

The EQ-5D and the 15D have been compared among patients groups in relation to diseases such as chronic obstructive pulmonary disease, epilepsy, rheumatoid arthritis, cancer, type II diabetes, HIV and rehabilitation patients with musculoskeletal, cardiovascular or psychosomatic disorders. Most of these studies have focused on outpatients (Stavem, 1999; Stavem et al., 2001; Linde et al., 2008; Stavem et al., 2005; Moock and Kohlmann, 2008; Kvam et al., 2011; Lillegraven et al., 2010; Kontodimopoulos et al., 2012). In addition to patient groups treated in ambulatory settings, the EQ-5D and the 15D have been used in comparisons of patients requiring inpatient care and in residents of a community (Hawthorne et al., 2001; Saarni et al., 2006; Saarni et al., 2010).

In general, the mean utility scores have been higher for the 15D than for the EQ-5D with the differences tending to be larger when the utility values are low. The difference in utility scores between instruments varies from 0.07 to at least 0.22. The EQ-5D has a tendency to a ceiling effect, i.e., showing a considerable concentration of scores at the maximum end of the scale (i.e., 1). The ceiling effect has varied from 10% in multiple myeloma patients to 42% in patients with epilepsy. In corresponding patient groups, a score of 1 was obtained by 0% and 14% of patients, respectively, when using the 15D (Stavem, 1999; Hawthorne et al., 2001; Stavem et al., 2001; Linde et al., 2008; Stavem et al., 2005; Saarni et al., 2006; Moock and Kohlmann, 2008; Saarni et al., 2010; Kvam et al., 2011; Lillegraven et al., 2012; Kontodimopoulos et al., 2012) (Table 2).

Table 2. Descriptive statistics of the utility scores using the EQ-5D and the 15D in different patient populations

	Mean utility score		Range		Ceiling effect (%)	
Patient group	**EQ-5D**	**15D**	**EQ-5D**	**15D**	**EQ-5D**	**15D**
Epilepsy[1]	0.81	0.88	-0.11-1.00	0.39-1.00	42.0	14.0
HIV/AIDS[2]	0.77	0.86	-0.33-1.00	0.43-1.00	29.0	10.0
Cancer[3]	0.74	0.86	---	---	29.0	7.0
Diabetes[3]	0.67	0.83	---	---	21.0	6.0
Heart failure[3]	0.59	0.77	---	---	8.0	1.0
Cardiovascular[4,5]	0.73	0.86	-0.07-1.00	0.56-1.00	21.6	5.7
Cardiovascular[6,5]	0.76	0.88	-0.07-1.00	0.55-1.00	26.1	10.2
Muskuloskeletal[4,5]	0.63	0.84	-0.18 -1.00	0.60-1.00	5.7	1.9
Muskuloskeletal[6,5]	0.67	0.87	-0.08-1.00	0.64-1.00	7.5	4.7
Psychosomatic[4,5]	0.57	0.76	-0.08-1.00	0.52-0.94	4.2	0
Psychosomatic[6,5]	0.57	0.79	-0.14-1.00	0.46-1.00	4.3	1.4

[1]Stavem et al., 2001, [2]Stavem et al, 2005, [3] Saarni et al., 2006, [4] Baseline HRQoL,
[5] Moock and Kohlman, 2008, [6]Follow-up HRQoL.

Among these relatively healthy patient populations, the EQ-5D produced HRQoL scores < 0, suggesting health states WTD. The percentage of patients scoring WTD has not been systematically reported, but the proportion of patients has varied from 3% to more than 6% (Moock and Kohlmann, 2008; Lillegraven et al., 2012). For these patients, the 15D score is positive, i.e., the 15D does not produce negative values.

The 15D has been shown to be more sensitive in detecting change in HRQoL and in discriminating different health states than the EQ-5D (Stavem, 1999; Moock and Kohlman, 2008; Saarni et al., 2010; Kontodimopoulos et al., 2012). This might be due to the notable ceiling effect of the EQ-5D. Furthermore, the richer health state descriptive system of the 15D may play a role. However, different results have also been reported. For example, in HIV patients, the responsiveness according to the clinical state did not differ between instruments, although the 15D showed a higher responsiveness to improvement (Stavem et al., 2005). In multiple myeloma and rheumatoid arthritis patients, the 15D did not detect a statistically significant change in the group of deteriorating patients, although the mean changes were negative (Linde et al., 2008; Kvam et al., 2011). In general, the quantity of the change in the HRQoL score is larger in the EQ-5D compared to the 15D (Stavem et al., 2001; Kvam et al., 2011).

2.2 Quality-adjusted life years

As mentioned above, a widely used approach for quantifying health gains is to use QALYs gained as a measure of the effectiveness of care. QALY combines two main outcomes of health care: mortality and morbidity, while also highlighting the populations' preferences (Bleichrodt and Johannesson, 1996). QALYs allow for comparisons between different patient populations and health-care interventions using a single, universal indicator (Prieto and Sacristan, 2003; Dolan et al., 2005; Brauer et al., 2006).

The foundation of QALY lies in utilitarian philosophy — people wish to maximise benefits and minimise harm (Dolan, 2001). An essential factor in the QALY model is the utility weight. The utility weight indicates the trade-off between the quality and the length of life (Scuffham et al., 2008). This trade-off means that individuals enjoying full health are unwilling to sacrifice any length of life; but, for individuals in, for example, a health state with a utility weight of 0.5, they are willing to sacrifice 50% of

their expected lifetime to become perfectly healthy for the rest of their lifetime. This implies that, for young people, low utility values mean greater losses than for older individuals. The utility weights lie along an interval scale; the same magnitude of change is equally valued across the entire scale (0–1). Thus, a change from 0.3 to 0.4 is as valuable as a change from 0.9 to 1 (Whitehead and Ali, 2010).

Although QALY is a widely accepted indicator to measure health gains, it has also been criticised. When applying TTO to the valuations of health states, one problem can be found in patients' reluctance to trade off lifetime (Nord et al., 2009). For example, when patients with advanced cancer valued their own health status using TTO, some respondents irrespective of their health state refused to trade off any lifetime. As a consequence, the utility index can be 1 (i.e., perfect health) even for patients with symptomatic, metastatic cancer (Perez et al., 2003). This problem might be overcome by using a generic, single-index HRQoL instrument which accounts for the symptoms experienced by the patient.

Another concern is the fairness of the maximisation of QALYs. Maximisation implies that health-care interventions should be focused on patients with the largest potential to benefit, while society may prefer to focus on patients who are worse off (National Centre for Priority Setting in Health Care 2008; Nord et al., 2009). However, the calculation of QALYs does not automatically imply the maximisation of QALYs and the principle for the allocation of health-care resources and prioritising treatments and patients may be different from the QALY maximisation.

In addition to the recommended generic HRQoL instruments, utility weights have been produced using holistic valuation methods and by mapping disease-specific measures to generic HRQoL instruments (Brauer et al., 2006; Marra et al., 2007; Kontodimopoulos et al., 2009). In addition, a utility catalogue exists which includes 2 159 different utility weights collated from previously published studies. However, utility weights are not available for all diseases and different utility weights for the same health states can be found depending upon which HRQoL instrument and valuation method was used. For example, the utility weights in myocardial infarction vary from 0.58 to 0.93 (Brauer et al., 2006).

2.2.1 The calculation of QALYs

Regardless of the criticism related to some aspects of QALYs, they are still considered a useful method for evaluating the effectiveness of different health-care interventions. Indeed, the United Kingdom's National Institute for Health and Care Excellence (NICE) regards QALYs gained as its principal measure of the outcome of care (Rawlins and Culyer 2004). The quantification of QALYs requires a decision about the duration of the benefit of care (time horizon of the calculation), the manner in which HRQoL changes during the time horizon and whether one is calculating the number of QALYs experienced or QALYs gained. At present, there is no consensus on how to tackle these issues.

To begin with, what is the most appropriate time horizon—i.e., the duration of the benefit of care—to be used in QALY calculations? For instance, guidelines from NICE advise calculating QALYs for an appropriate time horizon (Guide to the Methods of Technology Appraisal, 2008). The Finnish guidelines for the evaluation of medicines issued by the Pharmaceuticals Pricing Board of the Ministry of Social Affairs and Health state that the time period should be long enough to take into account all essential costs and health effects (Pharmaceuticals Pricing Board, 2011). As a consequence, diverse time horizons have been used varying from short time periods to tens of years. Some examples of time horizons include the follow-up time (Cuthbertson et. al., 2010; Kantola et al., 2010; Harris et al., 2011; Sultan and Hynes, 2011), life expectancy (Sznajder et al., 2001; Linko et al., 2010; Peek et al., 2010) and reduced life expectancy (Talmor et al., 2008; Mahonay et al., 2011; Malmivaara et al., 2011).

Due to the fact that the frequent measurement of HRQoL (e.g., on a daily basis) is normally not possible, attention should be paid to the manner in which HRQoL changes during the time horizon used

for analysis (Manca et al., 2005). Three different assumptions have been proposed: HRQoL changes linearly between measurement points; HRQoL remains constant from one measurement to the next and then changes overnight; and HRQoL changes at the midpoint between measurements (Billingham et al., 1999). In addition, a fourth assumption has been used within the critical care setting, namely, that the change in HRQoL takes place at the start of care (Karlsson et al., 2009).

When discussing QALYs, one must make the clear distinction between QALYs experienced and QALYs gained. This difference is illustrated using imaginary data as shown in Figure 2. Here, HRQoL has been measured at 1-year intervals in a hypothetical patient group without treatment. The resulting mean HRQoL scores (utility values) are shown on the vertical axis. HRQoL is assumed to change linearly between the measurement points and results in a curve. The entire area under the curve (AUC, the grey area) calculated using the trapezium rule represents the mean number of QALYs experienced by the patient group during the time horizon of four years, i.e., during their remaining life expectancy.

Had the patient group been treated, it would have experienced higher mean HRQoL scores and lived longer. Here, The AUC (grey and black areas) represents the mean number of QALYs experienced and the black area represents the mean number of QALYs gained by the patient group receiving treatment.

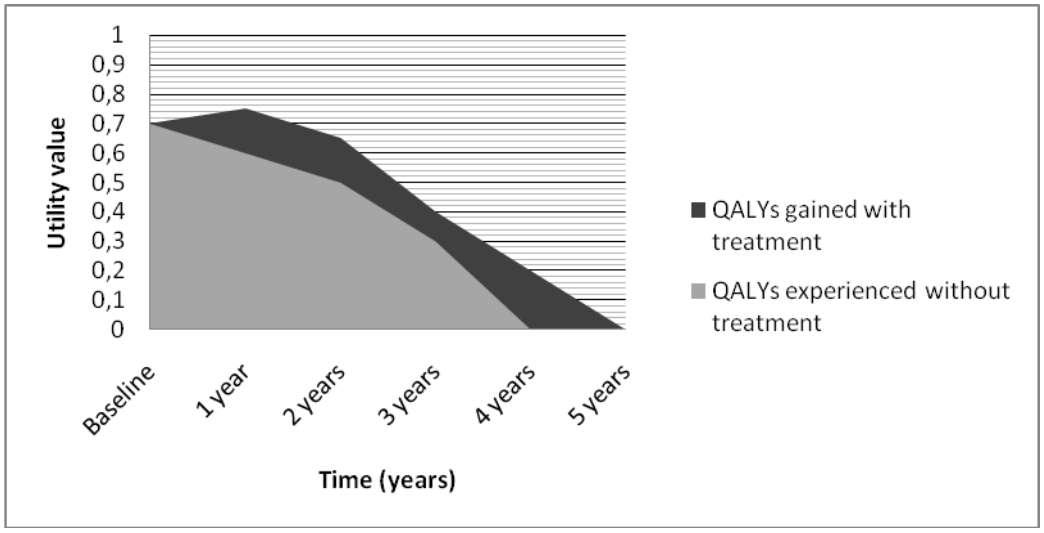

Figure 2. QALYs experienced and QALYs gained

When calculating QALYs, one should pay attention to the baseline utility value since it is strongly correlated with the number of QALYs (Manca et al., 2005). In circumstances where the baseline utility weight is unknown and it is challenging to obtain it such as in critical care setting, assumptions about the baseline utility weight must be made.

Although the calculation of QALYs includes several elements, in many cases the calculation methods are not explained transparently (Richardson and Manca, 2004; Schwappach and Boluarte, 2007; Rodriguez et al., 2011). In addition, the utility weights and the change in them are expressed as mean values and the dispersion of QALYs experienced or gained are not reported.

2.3 Critical care

Critical care is delivered in special units called intensive care units (ICU) or high-dependency units (HDU). The most serious conditions are treated and the most demanding forms of care are provided in ICU. Critical care is resource-intensive; medical staff is available around the clock and nurses can take care of only one to three patients at a time depending on the patients' states. In addition to the heavy personnel burdens, the critical care environment is technologically advanced featuring diverse equipment to monitor patients and to deliver demanding care such as mechanical ventilation and renal replacement therapy (Valentin and Ferdinande, 2011). From the patients' point of view, the critical care environment is stressful (Almerud et al., 2007). Typically, critical care patients are confined to bed, connected to monitoring devices via cables and are unable to express themselves. In addition to the physical discomfort, serious illnesses raise the fear of the discontinuance of life (Wang et al., 2008).

Although there is no rule regarding upon which patients ICU treatment should be focused, it is generally accepted that it should be focused on patients with reversible medical conditions with a high but not enormous risk of death (Task Force of the American College of Critical Care Medicine, 1999). In addition, it has been stated that admission to critical care requires that a patient's vital functions are threatened by an acute disease event, by surgical or other intensive treatment or when one or more of the vital functions have already failed and the patient needs demanding interventions. In addition to the life threatening condition, the patient should have the potential for recovery (Valentin and Ferdinande, 2011). The typical surgical treatments requiring critical care are, inter alia, cardiac surgery, neurosurgery and many arterial surgical procedures. Typical medical illnesses requiring critical care are acute myocardial infarction, cardiac arrest, respiratory failure, sepsis and neurological diseases such as stroke or cerebral haemorrhage (Mayer et al., 2000; Graf et al., 2005; Seferian and Afessa, 2006; Graf et al., 2008).

2.3.1 Critical care patients' survival

Although only patients with the potential to recover should be referred to critical care, the mortality rate among critical care patients is high at least within the first year after the initiation of treatment (Kaarlola et al., 2003; Rimachi et al., 2007; Karlsson et al., 2009; Linko et al., 2010; Khouli et al., 2011). Among general ICU patients, mortality has been reported to vary from 16% to 44% (Rotondi et al., 2002; Deja et al., 2006; Merlani et al., 2007; Cuthbertson et al., 2010) and hospital mortality from 24% to 58% (Graf et al., 2008; Khouli et al., 2011; Vaara et al., 2012) depending on the diagnostic category. For example, the 1-year mortality of sepsis patients was reported to be 41% (Karlsson et al., 2009), 46% for acute heart failure patients (Zannad et al., 2006) and 34% for patients with infections (Mayr et al., 2006).

The mortality rate is lower in ICU patients receiving elective surgery compared with emergency admissions (Niskanen et al., 1996). For instance, in cardiac surgery patients, the 6-month mortality has been reported to vary from 2% to 6% (Welsby et al., 2002; Schelling et al., 2003; Hein et al., 2006; Pätilä et al., 2006; Van den Heede et al., 2009). Long ICU stays predict higher mortality rates among cardiac surgery patients. Hospital mortality has been reported to vary from 8.5% to 52.9% after a prolonged ICU stay (Pappalardo et al., 2004; Gersbach et al., 2006; Hein et al., 2006; Gaudino et al., 2007).

In addition, long-term mortality is high among critically ill patients. For acute respiratory distress syndrome patients discharged alive from ICU, the 2-year mortality was 49%, while for general ICU patients alive 6 months after treatment at ICU, the 9-year mortality was 44% (Cheung et al., 2006; Stricker et al., 2011). For general ICU patients, the 2-year mortality including ICU mortality was 53% (Schenk et al., 2012), while for surgical ICU patients, the 6-year mortality was 54% (Timmers et al., 2011). For cardiac surgery patients, mortality varied according to the length of stay in ICU. In the group with a short ICU stay

(i.e., 3 days or less), the mortality rate during a 3-year follow-up was 9% compared to 34% in the group of long ICU stay patients (Hein et al., 2006). After an isolated aortic valve replacement (AVR) procedure, mortality during a 5-year follow-up was 11.4% for patients younger than 80 years and 28.1% for those 80 years or older (Saxena et al., 2012).

2.3.2 The costs of critical care

There are two notable issues to take into account in determining patient-specific costs. First, the cost per patient usually varies significantly, and second, the costs of treatment for a single patient can reach tens of thousands or even hundreds of thousands of Euros. For example, the costs per cardiac arrest patient were reported to vary from 1 708 € to 181 500 € (Graf et al., 2008), while those for general ICU patients ranged from 1 474 USD to 261 051 USD (Wachter et al., 1995). Low ICU costs usually indicate either a fast recovery or a fast death. For critical care patients, ICU costs and total hospital costs mostly depend on the length of the ICU stay (Graf et al., 2008; Niskanen et al, 2009; Linko et al., 2010).

The average total hospital costs for critical care patients have usually been reported to be on the order of 20 000 – 50 000 USD regardless of the reason for care, i.e., scheduled surgery or acute care. The tendency is that the care is more expensive in older patients (Agarwal et al., 2010; Gelsomino et al., 2011) and that patients at the highest risk are not necessarily the most expensive (Hamel et al., 2000). However, attention must be paid to the fact that the calculation and source of costs are not congruent across all studies (Table 3).

Table 3. Total hospital costs for cardiac surgery and general ICU patients in various studies (costs in USD unless otherwise stated)

Study	Patient group (N)	Mean costs (SD)	Median (Q1/Q3)	Cost source
Hamel et al., 2000	ARDS[6] Low-risk (292) Medium-risk (385) High-risk (286)	 59 096 (64 336) 70 130 (85 300) 59 310 (54 590)		Based on billing
Moran et al., 2004	ICU patients (1 333)	9 343[8] (-)		South Australian Health Commission Study
Dasta et al., 2005	Mechanical ventilation (18 590) Non-mechanical ventilation (32 419)	47 158 (57 703) 23 707 (34 545)		Based on billing
Graf et al., 2005	Medical ICU patients (190)	14 130[6] (-)		Patient-specific (variable) and non-patient-specific (fixed) costs
Cheung et al., 2006	ARDS (78)	128 860[7] (-)		Hospital Administrative Database
Karlsson et al., 2009	Sepsis (269)	32 563[5](-)		Estimated from total annual costs
Agarwal et al., 2010	CABG[1] 40-50 years (149) 50 – 60 years (605) 60-70 years (1006) 70-80 years (754) ≥ 80 years (268)	 27 580 (12 465) 30 904 (17 765) 33 758 (25 896) 37 426 (31 455) 42 115 (29 729)		Maryland Health Services Cost Review Commission
Bhamidipati et al., 2011	CABG[1] (1992) AVR[2] (352) MVR[3] (81) CABG+valve[4] (654)	35 017 (31 540) 37 759 (23 581) 47 274 (31 440) 44 965 (30 003)		Virginia Cardiac Surgery Quality Initiative Registry
Gelsomino et al., 2011	(70–79 years n 1230 ≥ 80 years n 1640) CABG[1] 70 – 79 years ≥ 80 years AVR[2] 70 – 79 years ≥ 80 years MVR[3] 70 – 79 years ≥ 80 years CABG+valve[4] 70 – 79 years ≥ 80 years		 12 212 (6 120/18 220) 16 759 (13 696/28 612) 14 431 (6 512/19 436) 19 760 (11 340/30 100) 16 019 (8 440/23 619) 21 307 (12 430/34 215) 16 650 (8 979/24 011) 22 666 (12 496/34 620)	Hospital and outpatient financial registers
Robinson, 2011	Cardiac valve replacement (37 hospitals)	43 733 (14 794)		Hospital finance department
Cohen et al., 2012	CABG[1] (870)	33 254 (9 782)		IMS Hospital Supply Index, statistical estimation
Iribarne et al., 2012	MVR (105) ≥ 75 years	60 289 (4 843)		Hospital finance department

[1]Coronary Artery Bypass Graft Surgery, [2]Aortic valve replacement, [3]Mitral valve replacement [4]Combined Coronary Artery Bypass Graft and valve Surgery, [5]€, [6]Acute respiratory distress syndrome, [7]Canadian $,[8]Australian$

2.3.3 HRQoL measurement in critical care patients

The HRQoL instruments most often used among acutely ill critical care patients include the SF-36 profile instrument and the EQ-5D. The 15D has also been used in a few studies (Elliot et al., 2004; Kantola et al., 2010). Due to the challenges in establishing baseline HRQoL in acutely ill critical care patients, various approaches have been used to calculate QALYs:

- Follow-up HRQoL has been compared to that for the general population (Kvale et Flaatten, 2003; Stricker et al., 2005; Deja et al., 2006; Ringdal et al., 2009; Linko et al., 2010; Orwelius et al., 2010; Timmers et al., 2011).
- The baseline HRQoL has been considered the value of 0 assuming that, without treatment, patients would die (Graf et al., 2005; Linko et al., 2010; Peek et al., 2010).
- The prevailing HRQoL value before treatment has been assessed using proxies (Badia et al., 2001; Wehler et al., 2003; Merlani et al., 2007; Hofhuis$_{s1}$ et al., 2008; Hofhuis$_{2}$ et al., 2008; Cuthbertson et al., 2010; Vaara et al., 2012) or by professionals (Kantola et al., 2010).

Studies comparing follow-up HRQoL among surviving patients to that for the population have usually found it to be impaired (Deja et al., 2006; Merlani et al., 2007; Karlsson et al., 2009; Cuthbertson et al., 2010; Linko et al., 2010). For example, the mean EQ-5D score 12 months after ICU treatment was 0.67 compared with the population's score of 0.82 (Cuthbertson et al., 2010). When the follow-up HRQoL was compared to the proxy-assessed baseline HRQoL measured using the EQ-5D, the median follow-up HRQoL was worse than the proxy-assessed baseline HRQoL in renal replacement therapy (RRT) patients (0.63 vs. 0.68) and similar in patients who did not need RRT (0.68 vs. 0.69) (Vaara et al., 2012). However, when the comparison was made to the HRQoL prevailing at the start of care, the follow-up HRQoL was clearly higher (0.30 vs. 0.70) in acute liver failure patients measured using the 15D (Kantola et al., 2010).

Furthermore, in cardiac surgery patients, the most commonly used HRQoL instrument has been the SF-36 profile instrument; however, in contrast to acutely ill critical care patients, the baseline HRQoL has been assessed by the patients themselves. The target has been to establish the HRQoL prevailing before treatment and, then, to compare it to the follow-up HRQoL after treatment (Schelling et al., 2003; Hawkes and Mortensen, 2006; Jokinen et al., 2010; Grady et al., 2011). The mean follow-up HRQoL scores after cardiac surgery have been found to significantly improve compared to the baseline HRQoL scores (Hawkes and Mortensen, 2006; Azzopardi and Lee, 2009; Loponen et al., 2009; Gjeilo et al., 2012; Markou et al., 2011). For example, the mean baseline HRQoL score measured using the EQ-5D in coronary artery bypass graft surgical (CABG) patients was 0.68 and 0.67 in AVR patients. At follow-up 12 months after treatment, the mean HRQoL scores were 0.78 and 0.71, respectively (Markou et al., 2011). Using the 15D, the mean baseline HRQoL score in CABG patients was 0.75 (Kattainen, 2004) and 0.83 (Loponen et al., 2009) and the follow-up score 6 months after treatment of 0.86 (Kattainen 2004, Loponen et al., 2009). Compared with the general population, mean HRQoL for cardiac surgery patients was reported to be fairly good (Gjeilo et al., 2006). However, although cardiac surgery patients on average benefit from the surgical procedure, 9% to 27% of patients experience a deterioration in their HRQoL score (Gersbach et al., 2006; Hawkes and Mortensen, 2006; Trouillet et al., 2011).

The reasons for these low and divergent follow-up HRQoL scores in some patients have been contemplated. The deterioration of HRQoL has not been associated with age, mortality risk, gender, duration of mechanical ventilation, sedation or length of ICU stay (Graf et al., 2005; Stricker et al., 2005; Deja et al., 2006; Gersbach et al., 2006; Hofhuis$_1$ et al., 2008; Davydow et al., 2009; Vest et al., 2011). By contrast, psychotic experiences, delusional memories, delirium and anxiety experienced during an ICU stay have been reported to affect follow-up HRQoL in a negative manner (Deja et al., 2006; Davydow et al., 2009; Loponen et al., 2008; Ringdal et al., 2009).

2.3.4 QALY calculation in the critical care setting

In principle, the measurement of QALYs is quite relevant in the critical care setting given that the goal of care is both to lengthen life and improve HRQoL in spite of the fact that the rule of rescue—that is, the duty to save an endangered life where a possible benefit can occur—applies. However, the methods used in QALY calculations vary between studies. For example, in some studies the cost per QALY gained has been calculated without using data based on a generic HRQoL instrument, although such data are an essential component of QALY calculations (Wu et al., 2007; Al-Ruzzeh et al. 2008; Graf et al., 2008; Yaghoubi et al., 2011; Gelsomini et al., 2011). Furthermore, the measurement of HRQoL has been reported in an incoherent way (Wu et al., 2007) and the population's age- and sex-matched HRQoL scores have been used among non-respondents (Linko et al., 2010). In addition, some studies comparing the effectiveness of two different forms of care have reported only the incremental cost per QALY ratio without separately reporting the costs and the number of QALY gained for the forms of care being compared (Eefting et al., 2003; Weintraub et al., 2004; Wu et al., 2007).

Furthermore, in some studies, the follow-up measurement of HRQoL has been performed before recovery can be expected to be complete, e.g., as early as 1 month after the surgical procedure (Eefting et al., 2003). In most studies, the number of QALYs has been expressed as an average, while some studies have used the sum of all QALYs gained (Karlsson et al., 2009). Additionally, the time horizon used for calculations has varied from 6 months to the entire life expectancy. Table 4 provides an overview of these studies.

Table 4: The reporting and calculation of QALYs in published studies

Study	Dg (n)	Area above/below the baseline HRQoL	Whole AUC	Time horizon in QALY calculation	Reported number of QALYs	Cost/QALY	HRQoL measurement
Sznajder et al., 2001	At least one organ failure (121 survivors)		x	Life expectancy according to the seriousness of the illness	3.9	4 110	EQ-5D
Eefting et al., 2003	OPCAB[2] (142)		x	12 months	0.790	14 188	EQ-5D
Weintraub et al., 2004	CABG (500)		x	12 months	0.694	12 831	EQ-5D
Sharples et al., 2006	Ventricular assist device (70)		x	Life time of transplantation patients	3.27	53 162	EQ-5D
Wu et al., 2007	AVR[4] (4 617) octogenarians and nonagenarians		x	Estimated lifetime	39 505	_____	Mapping quality of life scores to NYHA class I=0.85 II= 0.71 III =0.57 IV = 0.43
Al-Ruzzeh et al., 2008	CABG[1] (53) OPCAB[2] (44)	x		6 months	0.362 0.379	---------- ----------	Mapping WHOQOL-100 answers to the EQ-5D
Graf et al., 2008	Cardiac arrest (81 survivors)		x	Life time	1 766	13 805	Mapping SF-36 answers to the Health Status Index(HSI)
Karlsson et al., 2009	Sepsis (156)		x	Survival time or populations life expectancy	5 131	2 139	EQ-5D
Kantola et al., 2010	Acute liver failure (90)	x		3 years	1.44	64 732	15D
Linko et al., 2010	Acute respiratory failure (288 survivors)		x	Survival time or populations life expectancy	10 857	1 089	EQ-5D
Gelsomino et al., 2011	MVR[3] age $\geq$ 80 age $\leq$ 79		x	Estimated lifetime	_____	1 391 516	SF-36
Yaghoubi et al., 2011	Heart valve Homograft (30) Mechanical (30)		x	Follow-up time varying according to the day of discharge	1.8 1.13	1 253 2 629	SF-36

[1] Coronary Artery Bypass Graft Surgery, [2] Off-pump coronary artery bypass, [3]Mitral valve replacement, [4]Aortic valve replacement

2.4 Summary of the literature

Although the improvement of HRQoL and cost-effective care are important targets for health care, their measurement varies. First, one of the most often used HRQoL instruments—namely, the SF-36—is a profile instrument which does not readily allow for the calculation of the cost utility of care. Second, HRQoL scores measured using different generic HRQoL instruments are often regarded as universal, single-index scores, while the scores produced using these different instruments are not similar. Although attention has been paid to this problem in recent years, understanding of the applicability and comparability of different generic instruments is still unclear in the critical care setting. Possibly due to the 2002 Brussels Roundtable Consensus Meeting's recommendation, current knowledge stipulates that comparisons between different generic HRQoL instruments within critical care are, by and large, lacking, yet urgently needed. The responsiveness to change among various instruments—i.e., their ability to detect changes in HRQoL over time—needs to be studied in much greater detail than has been the case thus far, since a change in the HRQoL score is an indicator of the effect of care.

At present, there is no gold standard for the calculation of QALYs, which has led to the use of variable calculation methods, differences in the estimation of baseline HRQoLs in the critical care setting and varying time horizons. Moreover, inadequate attention has been paid to recording the patterns of recovery and the development of HRQoLs during a time horizon. In addition, understanding of the considerable effect the use of different generic HRQoL instruments has on the number of QALYs gained or experienced leaves much to be desired. Given that the demand for resource-intensive critical care is increasing, it is crucial to understand the effect of different HRQoL instruments, calculation methods and assumptions on the number of QALYs and the cost per QALY gained in the critical care setting.

3. Aims of the study

Treatment in the critical care setting is resource-intensive and likely to require even more resources in future due to increasingly demanding treatment modalities and the ageing of the population. Therefore, it is important to know the effectiveness and costs of different interventions. The overall aim of this series of studies was to identify factors causing differences and inaccuracies in the calculation of QALYs as a measure of effectiveness in the critical care setting. It is hoped that this will improve the quality and comparability of economic evaluations within the field.

The specific objectives were:

1. To compare the characteristics of two HRQoL instruments—the EQ-5D and the 15D—in the critical care setting. That is, are the HRQoL scores produced by the EQ-5D and the 15D interchangeable (study I).

2. To assess the sensitivity of the EQ-5D and the 15D in detecting a change in HRQoL, i.e., the responsiveness to change after treatment in the critical care setting. That is, which of the two instruments—the EQ-5D or the 15D—is more suitable for the evaluation of HRQoL in the critical care setting in terms of discriminatory power and responsiveness to change (study I).

3. To assess the effect of the HRQoL instrument used and the calculation method employed on the number of QALYs gained by treatment in the critical care setting. That is, what is the effect of the calculation method and the HRQoL instrument—the EQ-5D or the 15D—on the number of QALYs and the cost per QALY ratio (study II).

4. To estimate the excess or reduced mortality and lifetime gained or lost in patients treated in an ICU or HDU or after elective surgery. That is, how can the potential excess mortality within the critical care setting is taken into account in QALY calculations (study IV).

5. To evaluate the ability of routinely used predictors of operative mortality to also predict follow-up HRQoL and to assess the effect of patient characteristics and care-related factors on follow-up HRQoL. That is, can factors predicting mortality and morbidity be used to predict follow-up HRQoL in cardiac surgery patients (study III).

4. Patients and methods

4.1 Patients

The studies are based on two prospectively collected data sets of patients treated in an ICU, HDU or cardiac surgical intensive care unit (CSICU) at the Helsinki University Hospital. The follow-up time was 12 months in studies I and II, until death or 30 October 2012 in study III and 6 months in study IV.

The data in study I consisted of all patients treated in ICU or HDU between 1 January 2003 and 31 December 2004 (N = 3 600). They consisted of both acutely ill and electively treated critical care patients from all diagnostic groups of the International Classification of Diseases, 10th edition (ICD-10) except for the group of perinatal diseases. The most common distinct diagnoses were intoxication (T36, n = 213), peripheral atherosclerosis (I70.2, n = 196), cardiac arrest (I46.0, n = 179) and abdominal aortic aneurysm without rupture (I71.4, n = 106). The patient population was analysed as a whole and the data were gathered after an ordinary care process.

The data in study II is a subgroup of patients from study I and includes only those patients who received care on an emergency basis (N = 1 990). The care was deemed to have started on an emergency basis if the admission to both the hospital and ICU or HDU occurred on the same day. The largest diagnostic group was diseases of the circulatory system (n = 701). The most common distinct diagnoses were intoxication (T36, n = 210) and cardiac arrest (I46.0, n = 167). All other distinct diagnostic groups included distinctly fewer patients. The data were analysed as an aggregate.

The data in study III comprise two different patient populations. First, the data used in study I (n = 3 600) and, second, all patients treated in CSICU who returned a baseline or follow-up HRQoL questionnaire between 1 March and 31 December 31 2006 (n = 1 186). The total study population, thus, compromises of 4 786 patients. Cancer patients (n = 260), patients with an unknown diagnosis (n = 99) and patients from diagnostic groups that had less than 99 patients (n = 2 741) were excluded as inadequate for analysis. The most common diagnoses or surgical procedures were CABG (n = 498), aortic valve surgery (n = 253), intoxication (n = 230), cardiac arrest (n = 213), pneumonia (n = 207) and peripheral atherosclerosis (n = 201).

The data in study IV consisted of 980 consecutive, elective cardiac surgery patients treated in CSICU between 1 March 2006 and 31 December 2007. To be included in the study, the patients had to have waited for the operation for at least 7 days, indicating non-urgent surgery. The most common surgical procedures were conventional CABG (n = 333), AVR (n = 202), CABG and valve surgery (n = 169) and off-pump coronary artery bypass (OPCAB) (n = 118) (Table 5).

Direct health-care costs for the patients were obtained from the Ecomed® clinical patient administration system (Datawell Ltd., Finland), where all costs of treatment for individual patients in the hospital are routinely stored.

Table 5. Characteristics of the patient population

	Study I	**Study II**	**Study III**	**Study IV**
Unit	ICU[1], HDU[2]	ICU[1], HDU[2]	ICU[1], HDU[2], CSICU[3]	CSICU[3]
Number	3 600	1 990	2 741	980
Follow-up time	12 months	12 months	until death or until 30 October 2012	6 months
Male (%)	62.5	62.7	68.6	70.2
Mean age (years)	60.1	57.4	63.0	65.7
LOS[4] in critical care setting (median)	2.8	2.0	2.0	1.0
LOS[4] in hospital (median)	12.0	9.0	11.0	9.0
Costs (median, €)	16 106	12 090	15 738	16 103
Cost (range, €)	1 045-356 800	1 045-334 118	1 045-356 800	3 647-169 273

[1]Intensive care unit, [2] High dependency unit, [3] Cardiac surgical intensive care unit. [4]lenght of stay

Ethical approval for studies I, II and III was granted by the local Ethics Committee (§12/2002/4.2.2002). According to the ethics committee of the hospital, study IV did not require ethical approval because the study data was based on standard information gathered during the care process. Permission for the study was, thus, obtained from the administration of the Helsinki University Hospital (§69/28.05.2008).

4.2 Methods

4.2.1 HRQoL

HRQoL was measured using the 15D and EQ-5D HRQoL instruments 6 and 12 months after treatment in studies I and II. Baseline HRQoL was, thus, not assessed by patients. The first questionnaires with an accompanying letter and an informed consent form were sent to patients still alive 6 months after treatment and they were asked to return the questionnaires in a prepaid envelope via post. Follow-up questionnaires were sent to those patients who had returned the 6-month questionnaire and were still alive at 12 months. In the case of a nonresponse, one reminder was sent.

In contrast to study I, the data from study II also included deceased patients. For patients who died during the 1-year follow-up time, HRQoL was assumed to have changed to 0 at the moment of death.

Since the data did not include baseline HRQoL, it was estimated in two different ways in study II. Baseline HRQoL was either assumed to be 0—i.e., the patient would have died without treatment—or it was assessed retrospectively based on information obtained from patients' medical and nursing records and mapping the information onto the 15D and EQ-5D questionnaires by two health-care professionals. The proxy assessment was based on the status of the patient upon admission to ICU or HDU.

Altogether, the baseline HRQoL was assessed for 112 patients in different diagnostic groups according to ICD-10. Infrequent diagnoses were combined into the diagnostic group of "other diseases." The average for all baseline HRQoL assessments in each diagnostic group was used for analysis (Table 6).

Table 6. Mean proxy-assessed (by two health-care professionals) baseline HRQoL scores according to the 15D and the EQ-5D and based on information obtained from patient records

Diagnostic group	**EQ-5D mean (range)**	**15D mean (range)**	**Total number of patients**
Resuscitated patients	-0.594 (-0.594 - -0.166)	0.106 (0.106 - 0.106)	17
Neurological diseases	-0.508 (-0.594 - -0.166)	0.149 (0.106 - 0.398)	12
Respiratory organ diseases	-0.408 (-0.594 - -0.166)	0.253 (0.106- 0.575)	13
Intoxication	-0.420 (-0.594 - -0.430)	0.208 (0.106 - 0.459)	16
Infectious diseases	-0.465 (-0.594 - 0.280)	0.237 (0.106 - 0.587)	14
Gastrointestinal diseases	-0.379 (-0.594 - 0.587)	0.392 (0.106 - 0.654)	12
Other diseases	-0.248 (-0.594 - 0.710)	0.407 (0.106 - 0.739)	12
Vascular diseases	-0.339 (-0.594 - 0.002)	0.528 (0.240 - 0.755)	7
Heart diseases	0.089 (-0.594 - 0.710)	0.528 (0.106 - 0.810)	9
Average	-0.387 (-0.594 - 0.710)	0.285 (0.106 - 0.810)	112

In study IV, HRQoL was measured using the 15D once the patients were placed on the waiting list for surgery and 6 months post-operatively. In the case of a nonresponse, no reminders were sent (Table 7).

Table 7. Measurement of HRQoL

	Study I	**Study II**	**Study III**	**Study IV**
HRQoL instrument	EQ-5D, 15D	EQ-5D, 15D	None	15D
Baseline HRQoL	Not measured	Estimated	None	When placed on the waiting list
Follow-up measurement	At 6 and 12 months	At 6 and 12 months	None	At 6 months

The change in HRQoL was classified according to MCID in HRQoL (studies I and IV). The change was coded as negative if it was ≤ -0.08 using the EQ-5D and ≤ -0.03 using the 15D and positive if it was ≥0.08 using the EQ-5D and ≥0.03 using 1the 5D. Other values were coded as unchanged.

4.2.2 QALY calculation

Calculation of QALYs in study II was performed using four different calculation assumption sets (AS1–AS4). In Figure 3, the area under the curve—i.e., the grey area—depicts the number of QALYs gained using different calculation assumptions. Using calculation assumptions AS1 and AS2, the baseline HRQoL is assumed to be 0, while using calculation assumptions AS3 and AS4, the baseline HRQoL is the proxy-assessed HRQoL score reflecting the patient's state upon admission to ICU or HDU.

In AS1, HRQoL changes with treatment immediately to the level observed at the first follow-up point, t_1 ($HRQoL_{t1}$), after which it changes linearly to that observed at the second follow-up point, t_2 ($HRQoL_{t2}$). The theoretical maximum number of QALYs gained during 1 year is 1:

(AS1) $QALY = HRQoL_{t1} * D_1 + [(HRQoL_{t1} + HRQoL_{t2}) / 2] * D_2$,

where D_1 is the duration of follow-up from baseline to the first measurement and D_2 is the duration of follow-up from the first measurement to the second.

In AS2, HRQoL changes with treatment linearly during the entire follow-up time. The theoretical maximum number of QALYs gained during 1 year is 0.750:

(AS2) $QALY = (HRQoL_{t1} / 2) *D_1 + [(HRQoL_{t1} + HRQoL_{t2}) / 2] *D_2$.

In AS3, HRQoL changes with treatment immediately to the level observed at the first follow-up point, after which it changes linearly to that observed at the second follow-up point. The theoretical maximum number of QALYs gained during 1 year depends on the baseline HRQoL score:

(AS3) $QALY = HRQoL_{t1} * D_1 + [(HRQoL_{t1} + HRQoL_{t2}) / 2 * D_2] – (HRQoL_{t0} * D_{Follow\text{-}up})$,

where $D_{Follow\text{-}up}$ is the total duration of follow-up.

In AS4, HRQoL changes with treatment linearly during the entire follow-up time. The theoretical maximum number of QALYs gained during 1 year depends on the baseline HRQoL score:

(AS4) $QALY = [(HRQoL_{t0} + HRQoL_{t1}) / 2] * D_1 + [(HRQoL_{t1} + HRQoL_{t2}) / 2 * D_2] – (HRQoL_{t0} * D_{Follow\text{-}up})$.

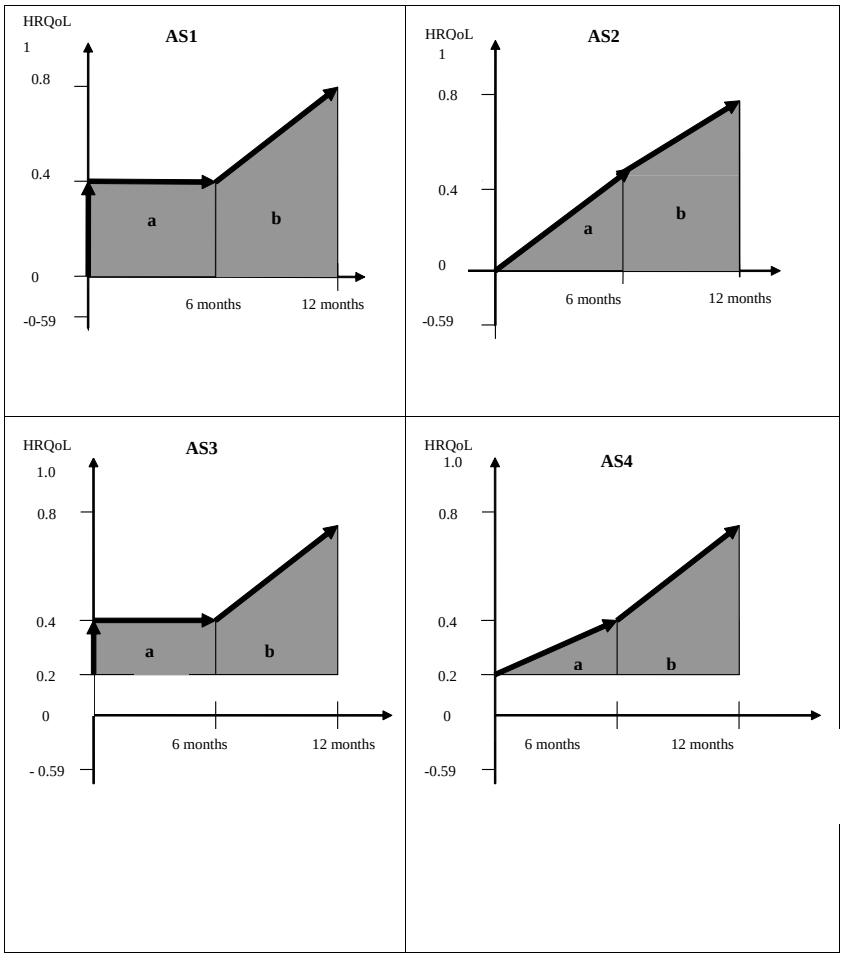

Figure 3. QALY calculation methods based on assumption sets 1–4.

4.2.3 Statistical methods

The descriptive statistics for continuous variables are reported as means, medians and ranges. The descriptive statistics for continuous variables with non-normal distributions are reported as medians and as percentages and ranges for categorical variables.

The association between the scores for the instruments was explored using the non-parametric Spearman rank correlation coefficient (Sarna (a), 2011), while agreement analysed using the Bland–Altman plot. The Bland–Altman plot is a graphical method used to compare agreement between two measurements by plotting the difference on the horizontal axis and the average of the measurements on the vertical axis. The reference lines show, respectively, the mean difference between the measurement and ±1.96 standard deviations from the mean difference (Bland and Altman, 1986).

Paired samples t-test was used to test the statistical significance of the differences in the mean number of QALYs obtained using the 15D and the EQ-5D. The Wilcoxon signed-rank test was used to test the statistical significance of the difference in the medians and distributions of the 15D and EQ-5D scores.

The discriminatory power of the instruments was explored by comparing the proportion of patients obtaining the ceiling score of 1 (ceiling effect) and the number of different health states. The agreement in the change in HRQoL scores (study I) and the direction of the change for the HRQoL scores between the baseline and 6-month measurements as observed in the data and predicted by the models (study

IV) were tested using the McNemar–Bowker test and the Cohen's kappa. The McNemar–Bowker test is a related samples test used when the measurement level is nominal and the variable includes more than two classes. It is used to test whether the matrixes of the measurements are symmetric. The Cohen's kappa is used with nominal variables to test the difference between observed agreement and random agreement. The Cohen's kappa varies from -1 to 1. A value of 1 means full agreement was found, 0 represents random agreement and a value of -1 indicates that the agreement is less than random indicating disagreement. A good agreement between measures requires a kappa value of at least 0.6 (Sarna (a), 2011).

A chi-squared test was used to determine whether the distribution of proxy-assessed patients across diagnostic groups deviated from that of all ICU patients across these groups. The Chi^2 test is an independent samples test to measure the differences between groups when the measurement level is nominal. The estimation is based on differences between observed and expected frequencies. The expected frequencies are calculated using probability mathematics (Ranta et al., 2005; Sarna (a), 2011).

The probability of death (study IV) was predicted using binary logistic regression using the maximum likelihood method. The binary logistic regression was used since the output variable was qualitative and had two possibilities: death occurs (1) or does not occur (0). The result of the analysis is the probability that the event—in our case, death—occurs in relation to not occurring. The maximum likelihood method generates coefficients which most likely produce the observations of the sample. The exponentiated coefficients are odds ratios, which express a percentage change in the probability of the event when the value of the explanatory variable changes 1 unit (Sarna (b), 2011).

The probability of death was explained in four stages. First, the patient characteristics before cardiac surgery, including a dummy value for coronary artery bypass grafting (CABG = 1, other heart surgery = 0), were entered in to the model. Second, a more parsimonious model including only those variables whose coefficients were statistically significant in the first stage was run. Third, in addition to the variables with statistically significant coefficients in stage one, the ICU-related variables were entered. Fourth, a more parsimonious model including only those variables whose coefficients were statistically significant in stages one and three was run.

The likelihood ratio test (LR-test) was used to determine whether model 2 produces a significantly better fit to the data than model 1 alone. The LR-test statistic is defined as (-2*Log-likelihood of model 1 + 2*Log-likelihood of model 2). The probability distribution of the test statistic is approximately a chi-squared distribution with df2 – df1 degrees of freedom, where df1 and df2 represent the number of free parameters from models 1 and 2, respectively. The LR-test is more appealing than the F-test when large samples are involved since it does not require an assumption of normality (Pindyck and Rubinfeld, 1991). For the same purpose, the correctness of classifications compared with the observed data was also analysed. The variance in HRQoL (the 15D score) at the 6-month follow-up among those who were alive at that time was explained using Tobit regression models by applying a similar four-stage approach with an LR-test to determine whether model 2 produces a significantly better fit to the data than model 1 alone. The Tobit regression was used because the follow-up 15D score is continuous but restricted, i.e., the maximum value is 1. The Tobit regression generates coefficients using the maximum likelihood method (Greene WH, 1998).

The relative survival method was used to estimate the possible excess mortality in different diagnostic groups in comparison to the population's mortality. The relative survival ratio (RSR) is calculated by dividing the observed interval-specific survival proportions of the patients by the expected ones in a comparable reference population. The expected survival proportions were derived from the mortality rates of the general population of Finland and stratified by sex, age and calendar time. We used RSR to resolve problems related to censoring caused by the limited follow-up times. The annual, bias-reduced relative survival method registers survival separately for each calendar year and survival time is estimated by using narrow age groups instead of the entire patient population. The annual, bias-reduced RSR method reveals the variation and disappearance of possible excess mortality during the follow-up time more easily than cumulatively reported survival. The narrow age groups dismiss the bias connected with informative

censoring, i.e., young patients' longer follow-up times and older patients' greater risk of death (Seppä and Hakulinen, 2009). For the prediction of the mean survival times, the cumulative observed survival proportions of the patients were estimated according to the follow-up time. Thereafter, the survival of the patients was extrapolated based on three different assumptions: 1) patients had the same mortality rates as comparable persons in the general population with respect to sex, age and calendar time, 2) patients had a 1% excess mortality and 3) patients had a 2% excess mortality during the rest of their lives.

A summary of the statistical methods used in the studies is given in Table 8.

Table 8. Statistical methods used by study and purpose

Study	Purpose	Method
I	To estimate the association between the HRQoL scores for the 15D and the EQ-5D	Spearman rank correlation
I	To estimate the agreement between the HRQoL scores of the 15D and EQ-5D	Bland-Altman plot
I & II	To test the statistical significance of the difference in the medians and distributions of the 15D and EQ-5D scores.	Wilcoxon signed-rank test
I & IV	To test whether the 15D and the EQ-5D yield a statistically significant similar result for the changes in HRQoL	McNemar-Bowker test
I & IV	To test the level of agreement in the direction of change of the 15D and the EQ-5D scores	Cohen's kappa
II	To test the statistical significance of the differences in the mean number of QALYs obtained using the 15D and the EQ-5D	Paired samples t-test
II	To test whether the distribution of proxy-assessed patients across diagnostic groups deviates statistically significantly from that for all ICU patients across groups	Chi^2 test
IV	To predict the probability of death by a certain point in time	Binary logistic regression
IV	To compare the fit to the data for two nested models	Likelihood ratio test
IV	To explain the variance in the follow-up HRQoL scores	Tobit regression
III	To estimate excess mortality	Relative survival ratio

The data were analysed using IBM SPSS Statistics (versions 17, 18 and 20), Limdep (version 7.0) and R-software. P-values of 0.05 were considered statistically significant.

4.2.4 Assessment of other parameters

Other parameters used in study III

Mortality was reported per 100 life years because the follow-up time varied between patients as the entry into the study occurred during several years. Mortality per 100 years was calculated by dividing the summed survival time by the observed mortality and multiplying the remainder by 100 years. The mortality rate was calculated separately for each diagnostic group and gender.

Clinical parameters used in study IV

The New York Heart Association classification (NYHA). NYHA is a classification system used to assess the stage of heart failure, ranging from class I to IV and is based on symptoms related to physical activity. Class I refers to no symptoms and no limitation, class II slight limitation and class III marked limitation in ordinary physical activity. Class IV means severe limitations even at rest (see http://www.abouthf.org/questions_stages.htm for further details).

The European method for cardiac operative risk evaluation (***EuroSCORE).*** Cardiac surgery patients' mortality risk was measured using EuroSCORE. EuroSCORE I was used, which was in use in the years 2006 and 2007. It consists of patient-related factors, the preoperative clinical state and cardiac-related factors. Patient-related factors include age, gender and co-morbidities such as diabetes mellitus, lung disease and impaired renal function. Cardiac-related factors include conditions such as recent myocardial infarction and left ventricular function. Surgery-related factors include the surgical procedure completed and the urgency of the procedure. The lowest possible score is 0, while a score of 6 or higher indicates a high risk level (see http://www.euroscore.org/calc.html for further details).

Sequential Organ Failure score (SOFA). The SOFA is used to assess the mortality risk in critically ill patients, and is based on the functioning of the respiratory, cardiovascular, hepatic, coagulation, renal and neurological systems. In it, each dimension is graded from 0 (normal) to 4 (most abnormal), where the total score can fall within the range of 0 to 24. The highest value of the SOFA was used in the analysis because it has been shown to correlate with 30-day mortality in cardiac surgery patients (Pätilä et al., 2006).

Body mass index (BMI). BMI was used to assess patients' nutritional status. It is calculated by dividing an individual's body mass (in kilograms) by the square of his/her height (in metres). Values under 18.50 indicate that an individual is underweight, while values above 25.00 indicate an individual is overweight (see http://apps.who.int/bmi/index.jsp?introPage=intro_3.html for further details). In critically ill patients, being underweight is associated with increased mortality (Garrouste-Orgeas et al., 2004).

The Therapeutic Intervention Scoring System (TISS-28). TISS-28 was used to measure the nursing workload during a patient's CSICU stay. The instrument consists of six categories: basic activities, ventilator support, cardiovascular support, renal support, neurological support, metabolic support and specific interventions. These categories are divided into 28 activities performed by nurses in ICU. Activities performed by nurses include, for example, dressing changes, care of drains and taking care of multiple vasoactive medications (Reis et al., 1996).

The Richmond Agitation and Sedation Scale (RASS). RASS was used to assess restlessness and deep sedation during a patient's CSICU stay. It is a 10-point scale describing a patient's state, ranging from alert and calm to combative or unarousable. The RASS scale ranges from -5 to 4, where negative values indicate sedation, 0 signifies that the patient is alert and calm and 1 and higher represent different levels of restlessness (Ely et al., 2003). Patients with RASS values of 1 or higher were classified as having experienced restlessness. Patients with a RASS value of -4 or -5 and an ICU stay longer than 2 days were classified as deeply sedated. RASS values of -4 and -5 during the first 2 days were ignored because, after a surgical procedure, patients are normally deeply sedated.

The Verbal Rating Scale (VRS). VRS was used to assess pain. The VRS scale ranges from 0 to 4, where 0 refers to no pain, 1 represents slight pain, 2 refers to moderate pain, 3 signifies severe pain and 4 indicates unbearable pain. VRS is also suitable for older individuals (Pesonen et al., 2008). For the analysis, patients were classified as having experienced severe or unbearable pain if their VRS scores were 3 or 4.

Nursing records. Structured, electronic nursing records were used to gather additional information about restlessness and the experience of pain. Patients were classified as restless if nursing reports included the words "disorientated", "confused" or "agitated" and as having experienced pain if the nursing reports indicated that the patient was in pain.

Other parameters. In addition, information was collected on complications (renal, neurological, respiratory, arrhythmia, urgent sternotomy and re-operation) that occurred during a patient's ICU stay and on the use of the intra-aortic balloon pump (IABP), the duration of ventilator treatment and the occurrence of nosocomial infections. The definition of a renal complication was based on the need for RRT or high doses of furosemide, and that of arrhythmias as arrhythmias other than atrial fibrillation and requiring medical or pacemaker interventions (Suojaranta-Ylinen et al., 2006). Delirium coinciding with a physical incident such as a brain infarction was classified as a neurological complication.

The clinical information in study IV was obtained from the Care Suite® clinical patient information system and the administrative database of the hospital. Information concerning patient care was gathered during the ordinary care process in all studies.

5. Results

5.1 Patients

The data from study I consisted of 929 patients who responded to both HRQoL questionnaires at the 6- and 12-month follow-up points. This represents 36% of the 2 600 patients alive at the time of the survey. Most of the patients were male (63%) and the mean (median) age was 60 (62) years old with a range of 16 to 92 years old. The median (mean) length of stay in the critical care setting was 2 (4.9) days ranging from less than 1 day to over 2 months. The median (mean) length of stay in hospital was 13 (18.7) days with a range of 1 day to 6 months. The median (mean) total cost for all treatment related to the illness episode was 17 871 (24 252) € ranging from about 2 000 € to almost 230 000 €.

Respondents and non-respondents differed statistically significantly regarding the mean age ($p < 0.001$), length of hospital stay ($p < 0.001$) and the cost of treatment ($p < 0.001$). Respondents were on average 4 years older (60 vs. 56), their median length of hospital stay was 1 day longer (13 vs. 12) and the median total costs for respondents was 2 138 € higher (17 871 € vs. 15 789 €).

The data from study II included patients whose treatment started on an emergency basis in ICU or HDU and who returned the completely filled HRQoL questionnaires both at 6 and 12 months or patients who had died before the first follow-up point (i.e., at 6 months). For patients who had died before the first follow-up point (n = 451), the HRQoL score was set to 0. In total, both follow-up HRQoL questionnaires were returned by 486 patients. Thus, the final data set comprises 937 patients, 47.1% of the original 1 990 eligible patients. Most of the patients were male (62%) and the mean (median) age was 61 (58) years old with a range of 16 to 98 years old. The median (mean) length of stay in the critical care setting was 3 (5.5) days ranging from less than 1 day to over 2 months. The median (mean) length of hospital stay was 9 (14.0) days ranging from 1 day to over 9 months. The median (mean) cost of the hospital stay was 14 392 (21 123) € ranging from about 1 000 € to over 330 000.

Respondents and non-respondents differed statistically significantly regarding the mean age ($p < 0.001$), length of hospital stay ($p < 0.001$) and total cost of treatment ($p < 0.001$). Respondents were on average 7 years older (61 vs. 54), for respondents the median length of stay in the critical care setting was 1.0 day longer (3.0 vs. 2.0), and the median total costs 1 954 € higher (14 392 € vs. 12 438 €).

The data from study III included patients treated in the critical care setting and who survived for more than 30 days (N = 2 445). Most of these patients were male (68.6%). For none of the diagnostic groups was the proportion of women clearly larger than that of men. The mean age for males was 62.4 years with a range of 16.4 to 89.4 years, while females had a mean age of 64.0 years ranging from 15.4 to 97.5 years. The median (mean) length of stay in the critical care setting was 2 (4.0) days ranging from less than 1 day to over 2 months. The median (mean) length of hospital stay was 10 (15.5) days ranging from 1 day to almost 6 months. The median (mean) cost of the hospital stay was 15 738 (22 559) € with a range of about 1 000 € to over 356 800 €. On average, the follow-up time was 5.9 years varying from 1 month to almost 10 years. The estimation of excess mortality is based on 14 381 person-years, which was 9 836 person-years for males and 4 545 for females.

The data from study IV consisted of cardiac surgery patients treated in CSICU, who waited for the scheduled surgery for 7 or more days and who answered both the baseline and 6-month follow-up HRQoL questionnaires or those who died during the follow-up time. The follow-up questionnaire was returned by 544 patients and 27 patients died during the follow-up. Thus, the final data set comprises 571 patients 58.3 % of the 980 eligible patients.

Most of the patients were male (70%) and the mean age was 66 years ranging from 21 to 90 years old. The median (mean) length of stay in CSICU was 1 day (2.8 days) ranging from less than 1 day to over 2 months. The median (mean) length of stay in hospital was 9 (11.5) days with a range of 2 days to more than 2 months. The median (mean) cost of the hospital stay was 15997 (20 978) € ranging from about 5 574 € to over 169 273 €.

The preoperative status of respondents and non-respondents differed statistically significantly only regarding the NYHA class (p = 0.002). Non-respondents more often fell into NYHA class IV than respondents (18.4% vs. 9.5%). With regards to the CSICU-related variables, respondents experienced severe or unbearable pain more often than non-respondents (p = 0.049). Among respondents, 22.6% experienced severe or unbearable pain compared to 17.2% of non-respondents (Table 9).

Table 9. Comparison of respondents and non-respondents in the studies

	Study I	**Study II**	**Study III**	**Study IV**
Patients (n)	929	937	2 445	571
Response rate (%)	36.0	47.1	Register-based study	58.3
Male (%)	63.0	62.0	68.6	70.1
Age (mean)	60.0	61.1	62.9	66.2
LOS[1] in critical care setting (median)	2	3	2	1
LOS[1] in hospital (median)	13	9	10	9
Costs (median)	17 871	13 251	15 738	15 998
Statistically significant difference compared to the non-respondents	Age*** LOS[1] in hospital*** Costs*	Age*** LOS[1] in critical care setting*** Costs*	--------	NYHA class** Severe or unbearable pain*

[1]Lenght of stay, p < 0.05* p < 0.01 **, p <0.001***

5.2 Agreement on HRQoL scores between the EQ-5D and the 15D

The instruments gave a different picture of patients' HRQoL. The HRQoL scores were uniform only in a minority of cases and neither of the instruments produced systematically higher or lower scores than the other. Although the ranking of the HRQoL scores (p < 0.001) was quite similar, the mean HRQoL scores (p < 0.001) and the distributions of the HRQoL scores (p < 0.001) differed in a statistically significant manner between instruments (study I).

The mean HRQoL was higher when assessed using the 15D compared to the EQ-5D. The mean HRQoL score was 0.832 at 6 months and 0.835 at 12 months when measured using the 15D and 0.731 and 0.735, respectively, when measured using the EQ-5D. The distribution of the HRQoL scores produced using the EQ-5D was discontinuous, had a long tail with low HRQoL scores and a peak with the highest possible HRQoL score. The long tail is partly explained by health states WTD, i.e., negative HRQoL scores. The number of negative HRQoL scores was about 3% using the EQ-5D. The distribution for the 15D was continuous, slightly skewed to the right and very low HRQoL scores were missing. The Bland–Altman graphical method verified the differences between the instruments.

The dissimilarities between the scores for the instruments were particularly evident at both ends of the measurement scales (Figure 4).

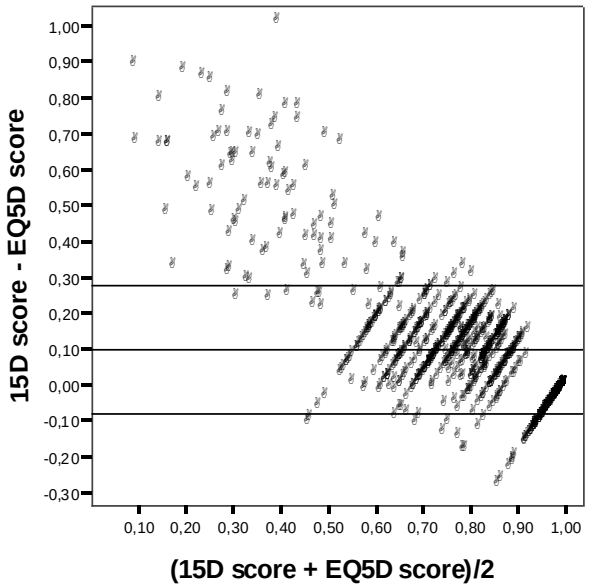

Figure 4. Agreement between the 15D and the EQ-5D scores at 6 months illustrated using a Bland–Altman plot (the horizontal lines represent the mean difference between the scores for the instruments ± 1.96 SD of the mean).

5.3 Comparison of the discriminatory power and responsiveness between the EQ-5D and the 15D

The discriminatory power was better with the 15D than with the EQ-5D. The ceiling effect—i.e., an HRQoL score equal to 1—occurred more infrequently using the 15D than with the EQ-5D. Using the 15D, 5.8% of patients at 6 months and 7.8% of patients at 12 months after treatment scored at the ceiling. The corresponding figures for the EQ-5D were 26.2% and 28.6%, respectively. Furthermore, the 15D was more sensitive in distinguishing between different health states than the EQ-5D. Using the 15D, the number of different health states was 767 at 6 months and 745 at 12 months. The corresponding figures for the EQ-5D were 79 and 70 (study I), respectively.

Although on average there was no change in the mean HRQoL score between 6 and 12 months, changes were observed at the individual level. The magnitude of change in the HRQoL scores was almost three-fold using the EQ-5D compared to the 15D. Using the EQ-5D, the magnitude of change varied from -0.82 to 0.88, while with the 15D, the range was -0.33 to 0.32. A clinically important change (equal or more than MCID) in the HRQoL scores was detected more often using the 15D than the EQ-5D. Using the 15D, 46% of patients were in the group of unchanged HRQoL, whereas the corresponding figure for the EQ-5D was 61%. Thus, the 15D was more sensitive to detecting a clinically important change, but the quantity of change was clearly smaller compared to that seen when using the EQ-5D.

The instruments yielded a different picture for the clinically important change in the HRQoL score ($p < 0.001$). HRQoL measured using the 15D improved for 29.3% of patients. Among those patients, the EQ-5D classified 1.6% into the group of deteriorated HRQoL and 16.8% into the group of unchanged HRQoL. The classification was consistent between instruments in 10.9% of all cases in the group of improved HRQoL according to the 15D. As stated above, the EQ-5D classified 61.0% of the patients into the group of unchanged HRQoL. Among those patients, the 15D classified 11.8% as belonging to the group of deteriorated HRQoL and 16.8% as falling into the group of improved HRQoL. There was agreement

between the instruments regarding the direction of change in 53.4% of cases; consequently, the Cohen's kappa was low 0.247 ($p < 0.001$) indicating a fair degree of agreement between the instruments (study I).

5.4 The effect of the HRQoL instrument on the number of QALYs

The 15D and EQ-5D instruments performed differently when using various calculation assumptions. The most significant differences between the instruments were detected when applying calculation assumptions AS3 and AS4, which both used proxy-assessed baseline HRQoL scores. Applying AS3 and AS4, the 15D generated the smallest and the EQ-5D generated the largest number of QALYs. Using AS1 and AS2, in which the baseline HRQoL was assumed to equal 0, the number of QALYs was quite comparable to the distinction that a QALY loss occurred only when using the EQ-5D. In that case, the QALY loss indicates the existence of health states WTD. Consequently, a QALY loss using the 15D occurred only when the proxy-assessed baseline HRQoL scores were used (AS3 and AS4).

In addition, the maximum number of QALYs gained revealed dissimilarities between the instruments. The maximum number of QALYs gained during a year was 1 QALY when using the 15D and 1.6 QALYs when using the EQ-5D. The EQ-5D produced more than 1 QALY a year when calculation assumptions AS3 and AS4 (using the proxy-assessed baseline HRQoL score) were employed. The percentage of cases in which more than 1 QALY per year was found reached 30.5% under AS3 and 7.4% under AS4. Furthermore, when using calculation assumptions AS1 and AS2 (the baseline HRQoL was 0), the way of recovery had an effect on the maximum number of QALYs gained. Applying AS1, the maximum QALY gain was 1 and 0.750 when applying AS2 irrespective of the instrument used (study II).

5.5 The effect of the calculation method on the number of QALYs and the cost per QALY ratio

The calculation method used (AS1–AS2) had a clear effect on the number of QALYs ($p < 0.001$) irrespective of the instrument used (the 15D or the EQ-5D). The mean number of QALYs in 1 year varied from 0.178 (AS4) to 0.419 (AS1) using the 15D and from 0.275 (AS2) to 0.550 (AS3) using the EQ-5D. As a consequence, the cost per QALY ratio varied according to the calculation assumption employed, i.e., according to the baseline HRQoL and the way in which recovery was assumed to have taken place. The cost per QALY ratio for the 15D was the lowest (50 412 €) when applying AS1 (the baseline HRQoL was 0 and the recovery took place immediately) and for the EQ-5D (38 405 €) when applying AS3 (the baseline HRQoL was proxy-assessed and the recovery took place immediately). The highest cost per QALY ratio for the 15D (118 688 €) was found when applying AS4 and for the EQ-5D (76 811 €) when applying AS2 (study II) (Table 10).

Table 10. The cost per QALY ratios for the 15D and the EQ-5D applying different calculation assumptions (cost in €)

	Cost/QALY ratio		Difference between the instruments	
	15D	EQ-5D	(€)	(%)
AS1	50 412	57 713	-7 301	14.5
AS2	67 271	76 811	-9 540	14.2
AS3	90 657	38 405	52 252	57.6
AS4	118 668	51 269	67 399	56.8

5.6 The effect of excess mortality and follow-up time on the extrapolated life expectancy

The relative survival method disclosed the excess mortality in critical care patients compared with most of the cardiac surgical patients. The excess mortality differed between diagnostic groups and varied over time without a predictable trend. The most pronounced excess mortality was seen in gastric ulcer patients, which varied from 5% to 16%, and in pneumonia patients, which varied from 2% to 13%. Another extreme was off-pump patients, who in comparison to the population were at least to some degree spared from death.

In extrapolating lifetime, the duration of the follow-up time had an effect on life expectancy. In diagnostic groups with excess mortality, the use of short follow-up times—e.g., 3–5 years instead of 9 years—gave a longer life expectancy than longer follow-up times (study IV).

5.7 The ability of the indicators predicting mortality to predict follow-up HRQoL

Of the variables describing the preoperative state including the dummy variable for CABG versus other heart surgeries performed, EuroSCORE predicted the probability of death during the 6-month follow-up (model 1) in a statistically significant manner. Of the ICU-related variables, renal, respiratory and neurological complications as well as urgent sternotomy turned out to be additional significant explanatory factors (model 2). The LR-test indicated that the fit of model 2 to the data was significantly better than that of model 1 alone (chi-square = 67.259, df = 5, $p < 0.001$). Model 1 was not able to correctly predict a patient as being dead (0% correct, total percentage correct 95.9%), whereas model 2 correctly predicted 99.3% of patients as being alive, and correctly predicted 30.4% as being dead (total percentage correct, 96.4%).

Of the variables describing the pre-operative state including the dummy variable for CABG vs. other heart surgeries performed, being male, having diabetes mellitus and the baseline 15D score explained the post-operative variance of HRQoL at the 6-month follow-up (model 1) in a statistically significant manner. ICU-related variables, severe or unbearable pain and restlessness during treatment in ICU were found to be additional statistically significant explanatory factors (model 2). The LR-test indicated that the fit of model 2 to the data was significantly better than that of model 1 alone (chi-square = 25.622, df = 5, $p < 0.001$).

Regarding the clinically important change in the HRQoL scores between the baseline and 6-month follow-up measurements, the observed data and the predictions from model 1 had the same direction in 63.4% of patients. Overall, the observed data and the predictions using model 1 resulted in a different

picture for the changes (McNemar–Bowker = 29.629, $p < 0.001$) and the agreement between them was poor (kappa = 0.264, $p < 0.001$). The observed data and the predictions from model 2 had the same direction in 63.3% of patients. The observed data and the predictions from model 2 resulted in a different picture for the changes (McNemar–Bowker = 22.612, $p < 0.001$) and the agreement between them was poor (kappa = 0.266, $p < 0.001$). The predictions using models 1 and 2 resulted in a similar picture for the changes (McNemar–Bowker = 1.377, $p = 0.711$) and the agreement between them was quite good (kappa = 0.690, $p < 0.001$). Thus, model 2 was not better than model 1 in predicting the clinically important changes in HRQoL.

Among the scheduled cardiac surgery patients alive at the 6-month follow-up point, 22.6% had experienced severe or unbearable pain and 14.4% experienced restlessness. For patients who had experienced severe or unbearable pain, the mean follow-up HRQoL score was 0.844 (SD ± 0.10) and 0.823 (SD ± 0.123) for patients who had experienced restlessness. For patients, who were free from both symptoms, the mean follow-up HRQoL score was 0.886 (SD ± 0.098) (study III).

6. Discussion

6.1 Main results

Our findings corroborate the assertion that the number of QALYs is not a universal measurement, but depends on how the factors taken into account in QALY calculation are chosen and defined. Thus, an acceptable threshold for the cost per QALY ratio is difficult to establish as long as QALY calculation methods are not standardised. Additionally, the costing methodology needs to be standardised. The 15D showed more discriminatory power and responsiveness to change than the EQ-5D in the critical care setting. Furthermore, the 15D performed more consistently using different calculation methods than did the EQ-5D. The inconsistency of the EQ-5D in QALY calculations is caused by the negative HRQoL scores it can produce. To allow for comparisons between the results from different studies, agreement is needed on how the baseline HRQoL score in studies from critical care setting is defined.

6.2 The discriminatory power and responsiveness to change of the EQ-5D and the 15D

The better discriminatory power and responsiveness to change of the 15D is most likely explained by the richer health state descriptive system in comparison to that of the EQ-5D. For example, ventilator treatment is a dominant form of care in the critical care setting, yet the EQ-5D does not include a dimension concerning breathing. Since patients treated in the critical care setting represent a wide spectrum of medical specialties, one would assume that a more comprehensive HRQoL instrument than the EQ-5D would be more appropriate for describing the health states of the patients cross-sectionally as well as changes in them over time (study I). The discriminatory power of the 15D is superior to that of the EQ-5D in this patient group. This has also been shown to be true in many other patient groups. The lower discriminatory power of the EQ-5D is at least partly explained by the pronounced ceiling effect (Stavem et al., 2001; Stavem et al., 2005; Kattainen et al., 2005; Saarni et al., 2006; Moock and Kohlmann, 2008; Färkkilä et al., 2013; Torvinen et al., 2013). This was also the case even in the patient group studied in which one would not a priori expect such a high percentage with a score of 1 (“full health”) as the group which was treated in ICU or HDU for serious, even life-threatening conditions just 6 months earlier.

The distribution of HRQoL scores differed between the instruments suggesting that reporting the mean value alone may not provide an adequate picture of the patients’ HRQoL. Consequently, in addition to the mean values, the distribution of HRQoL scores should be reported. It is also evident that the distribution of the EQ-5D scores is usually such (e.g., discontinuous, two- or three-peaked or with a high ceiling effect) that conventional statistical methods are not suitable for data analysis.

The EQ-5D was not sensitive in detecting clinically important changes and most of the patients were in the group of unchanged HRQoL. The range of changes varied from -0.82 to 0.88 with an average of 0.005. However, when a clinically important change occurred, the change in the overall score was large. This is a direct result of the health state descriptive system and the UK TTO valuation algorithm: due to the inclusion of only three levels per dimension, the distances between the levels are value-wise quite great. If the instrument is able to detect a change from one level to another, the change in the overall score is automatically quite large. This phenomenon is accentuated if a change takes place to or from level 3 on any dimension — then, the score changes by an additional 0.269 points in either direction, respectively, due to the N3 term (Dolan, 1997).

The 15D was sensitive in detecting clinically important changes, but the mean change in the HRQoL score remained modest. The range of changes varied from -0.33 to 0.32 with an average of 0.003. Thus, for both the 15D and the EQ-5D, the mean change was almost 0 and reveals nothing about the important difference in the distribution of patients between those whose HRQoL improved, remained

unchanged or deteriorated. In this respect, the instruments yield a clearly different picture and the agreement between them is only fair. Reporting the distributions for changes in scores is quite informative; unfortunately, thus far, this has not been a common way of reporting results. It would, however, promote the allocation of health-care resources to patients who can benefit from treatment. Such information could also help to develop care processes.

6.3 Calculating QALYs

According to our results, in addition to the instrument used, the calculation assumptions concerning the baseline HRQoL, the path to recovery and whether QALYs experienced (the entire area under the curve) or gained (the difference between the baseline and follow-up HRQoL) were assessed greatly influenced the number of QALYs. It is unfortunate that studies reporting the effectiveness of care do not normally clearly state the calculation methods used in quantifying QALYs (Richardson and Manca, 2004; Schwappach and Boluarte, 2007; Rodriguez et al., 2011) despite the fact that comparisons of the results from different studies and making conclusions based on them require transparent reporting.

The difference between various calculation methods—i.e., whether QALYs experienced or gained are explored—is clear when using HRQoL instruments operating with positive values such as the 15D. The number of QALYs experienced is higher compared to the number of QALYs gained. In our studies the average number of QALYs was almost twofold when QALYs experienced were assessed compared to QALYs gained. The issue, however, is different when HRQoL instruments producing negative scores are used. In the material used here, the EQ-5D produced fewer QALYs when QALYs experienced were assessed compared to those gained.

When reporting QALYs experienced, one must bear in mind that the effect of care is rarely directed to all dimensions of HRQoL and that the follow-up HRQoL is affected by factors prevailing before treatment, i.e., the baseline HRQoL. Assessing QALYs gained more accurately combines the observed change in the HRQoL score to the care delivered and also reflects society's possible preference for allocating resources to the treatment of worse-off patients (Nord et al., 2009). Given that the baseline HRQoL has such a significant effect on the follow-up HRQoL, consensus on how to define the baseline HRQoL is highly desirable. It is likely that those acutely ill patients for whom the need for intensive care is the consequence of imminent organ failure can assess their HRQoL independently for themselves. The question on how to define baseline HRQoL is, thus, relevant only in patients already experiencing organ failure.

In addition to the problem of how to quantify the baseline HRQoL, the calculation of QALYs experienced or gained is affected by assumptions concerning recovery. If the recovery process is assumed to take place immediately and the baseline HRQoL is assumed to be 0, the calculation of QALYs experienced makes it possible to experience 1 QALY within a year, whereas if the recovery is assumed to take place linearly, the maximum number of QALYs experienced can only be 0.750. Achieving the latter assumes that the measurement of HRQoL takes place at 6 and 12 months after the baseline measurement.

The reasoning regarding how recovery takes place is often lacking in studies reporting on the effectiveness of care. The assumption most often used within critical care—i.e., that recovery materialises at the start of care—is presumably too optimistic. It is evident that the way in which and the time during which recovery takes place vary between different diseases and, thus, the calculation method used should reflect reality. It is evident that the follow-up time points for assessing HRQoL should be tailored, if possible, to mimic the expected pattern of recovery for individual diseases. Nevertheless, the approach used for defining the pace of recovery should be reported clearly and openly.

6.4 Baseline HRQoL in the critical care setting

According to previous literature and our results, the baseline HRQoL predicts the follow-up HRQoL. Consequently, assumptions about the baseline HRQoL have significance for the calculation of the quantity of QALYs. Within critical care, the approach used for the determination of baseline HRQoL is, thus, of vital importance.

Because of the difficulties in obtaining baseline HRQoL in the critical care setting, the baseline HRQoL has sometimes been ignored and only the follow-up HRQoL has been used in the evaluation. In such cases, it has usually been compared to the population's HRQoL (Kvale et al., 2003; Stricker et al., 2005; Deja et al., 2006; Ringdal et al., 2009; Linko et al., 2010; Orwelius et al., 2010; Timmers et. al 2011). The population, however, is probably not an optimal source for comparison since HRQoL for healthy people may be better than that for individuals after a serious illness. As a consequence, the use of the population's HRQoL as a comparison might provide an estimate which is too pessimistic regarding the effectiveness of critical care.

Another assumption—namely, that without treatment the patient would die and, consequently, the baseline HRQoL is 0 (Graf et al., 2005; Linko et al., 2010; Peek GJ et al., 2009)—is also inaccurate. First, the result is the same irrespective of whether QALYs experienced or gained are calculated. Second, the entire scale of the instrument extending to the negative side is not in use. If health states WTD are considered possible, one would expect that these occur primarily at the beginning of intensive care when the acutely ill patient is often not able to breathe spontaneously, is unable to express her-/himself and is totally dependent on her/his caregivers. Third, critical care is directed to patients with the potential to recover, not to those dead. Thus, the initial value of 0—i.e., equivalent to being dead—is inappropriate. Furthermore, patients requiring intensive care for different reasons must be in a comparable position regarding the assessment of the effectiveness of care.

Additionally, the assessment of baseline HRQoL by professionals at the start of care is problematic, especially in relation to the assessment of mental dimensions such as depression and dimensions concerning patient experiences such as pain. However, some dimensions, such as moving, breathing and usual activities, can certainly be assessed by proxy quite accurately. Despite its drawbacks, the assessment of baseline HRQoL by proxy may better reflect reality than other alternatives. Consensus regarding the assessment of mental dimensions and dimensions concerning patients' experiences, however, is needed in order to enable robust comparisons between different HRQoL and cost-utility studies.

HRQoL prevailing months before critical care assessed by proxy has not been used in QALY calculations, although it has been measured in some studies. One reason for this might be that it has been higher than the follow-up HRQoL, suggesting that there is no benefit from receiving critical care. It is likely that the HRQoL prevailing months before critical care does not necessarily reflect the health state prevailing when critical care starts. This applies to all groups of critical care patients, i.e., acutely ill patients with and without failed vital functions as well as scheduled surgical patients. The assessment of the baseline HRQoL in the critical care setting may require different presumptions depending on the specific patient population.

6.5 Problems concerning negative HRQoL scores

In relation to the theory of QALY according to which 1 QALY represents a year in perfect health, negative HRQoL scores cause problems. Due to negative HRQoL scores, the number of QALYs gained or lost within a year can be larger than 1 QALY, thus dismissing the rationale for QALYs. When using the EQ-5D with the UK TTO valuation algorithm, the maximum number of QALYs gained is 1.6

during a year since the scale of the EQ-5D extends from -0.594 to 1. In addition, a change from a negative HRQoL score to death implies an improved HRQoL and, consequently, a QALY gain. Furthermore, a decreasing negative HRQoL score between the baseline and follow-up—e.g., from -0.3 to -0.1—indicates a QALY gain despite the patient remaining in a state WTD.

In the material here, which is based on the EQ-5D results applying the UK TTO valuation algorithm, about 3% of patients treated in ICU or HDU but living at home and capable of answering the questionnaires obtained a HRQoL score WTD. Does this really mean that those patients considered death a more desirable alternative than continuing to live in their present health state? We do not know, since to our knowledge no one has ever directly asked such patients whether they agree that their health state is WTD and whether they would rather die than go on living in their present health state. It is difficult to envision what useful information varying negative scores—e.g., -0.594 vs. -0.040—carry and the practical implications they might convey.

Of course, it is also possible that a positive HRQoL score would result for someone who would regard her/his health as WTD with instruments generating only positive scores like 15D. Technically, it would be easy to scale the scores so that negative scores are also allowed. But, the question becomes: how does one establish a reasonable, non-arbitrary, lowest possible negative score?

Remaining on the positive side of the scale is consistent with the ethical climate in most societies: even if the individual considers her/his health state WTD and even if most people would regard it as such, legislation does not acknowledge such states and allow people in those states to be helped to die for improving their quality of life. Thus, most societies assign a positive score to all health states (except, perhaps, for brain death) in their health policy. From this point of view, nothing is gained by allowing negative scores, while considerable analytical and ethical complexities are created in doing so (Sintonen, 1995).

6.6 The time horizon used in QALY calculations

According to previous literature and the results presented here, mortality among critical care patients is elevated (Hein et al., 2006; Stricker et al., 2011; Timmers et al., 2011; Shenk et al., 2012) indicating that the time horizon used in QALY calculations should be decided upon carefully. At present, different time horizons are used and the seriousness of the illness is not necessarily taken into account (Weintarub et al., 2003; Al-Ruzzeh et al., 2008; Graf et al., 2008; Linko et al., 2010).

The use of the follow-up time as a time horizon might result in inaccuracies due to censoring and possible long-lasting excess mortality. An additional disadvantage is that the possible complications of the demanding care process, which may last for the entirety of the remaining lifespan, are accounted for in the calculations for a limited period of time only. In principle, one purpose of critical care is to save lives, which means that the effect of care lasts for the duration of a patient's lifetime. Consequently, using life expectancy as a time horizon within the critical care setting is a better alternative than the follow-up time. The Finnish guidelines for the evaluation of medicines state that the time horizon should be long enough to take into account all essential costs and health effects (Pharmaceuticals Pricing Board, 2011).

The annual RSR makes it possible to decipher the potential annual excess mortality compared to an age- and sex-standardised population-specific mortality. According to the results here, general critical care patients experienced excess mortality during the entire follow-up period; but, in scheduled cardiac surgery patients, the mortality was almost comparable to the population's mortality. Consequently, the excess mortality of patients in intensive care is dependent on the form of care and should be assessed according to diagnostic group or form of care.

The estimation of the life expectancy for patients using the bias-reduced method—i.e., in narrow age groups—turned out to be useful, since it was possible to estimate the age-specific reduction in

lifetime in patient groups with excess mortality. This is clearly a methodological improvement adding to the credibility of the estimation of life expectancy compared to previous methods used to adjust for excess mortality, where a fixed number or proportion of life years was reduced from the general population's life expectancy regardless of age (Malmivaara et al., 2011; Talmor et al., 2008). In addition, this made it possible to overcome the use of concurrent reductions in lifetime, the relevance of which from the point of view of life expectancy is dependent on age.

The approach used here also allows for the extrapolation of life expectancy using different mortality rates depending on whether the mortality rate after the follow-up time is equal to that of the population or whether varying degrees of excess mortality remain for the duration of one's lifetime. This is an important advantage since excess mortality does not necessarily end at the same time as follow-up and may continue for unspecified time periods. In addition, the length of the follow-up time had an effect on the extrapolated life expectancy; thus, sensitivity analyses are particularly important in illnesses with high mortality rates when follow-up times are limited.

In addition to illustrating excess mortality, the annual RSR can also reveal reduced mortality. For example, in ageing populations evaluated as fit enough to undergo demanding forms of care, the age-standardised population's life expectancy might not be similar to that of the patients. Furthermore, in an ageing population, the mortality rate is naturally high and the reported high mortality rate can be incorrectly associated with the treatment under evaluation even though it is a consequence of the patients' age.

6.7 Factors affecting the follow-up HRQoL

The probability of death and the follow-up HRQoL were predicted by different factors. None of the CSICU-related morbidity or mortality factors, such as renal or respiratory complications, predicted the follow-up HRQoL. Given that a good HRQoL is an important objective of health care, its evaluation demands HRQoL-specific indicators. Previous research (Deja et al., 2006; Davydow et al. 2009; Loponen et al., 2008; Ringdal et al., 2009) has indicated that variables predicting the mortality risk are not valuable in the prediction of follow-up HRQoL within the critical care setting. Instead, patients' subjective experiences during treatment have been found to be important from the point of view of the follow-up HRQoL.

When calculating the number of QALYs as a measure of the effectiveness of care, it is important to identify factors affecting the follow-up HRQoL. The existence of such factors might enhance or dilute the actual effect of the care delivered, i.e., follow-up HRQoL scores are improved or impaired due to other factors than the care delivered. The study reported on here suggests that restlessness experienced during an ICU stay has a detrimental effect on the follow-up HRQoL. This result is consistent with those from earlier studies that have reported the negative effect of restlessness during a hospital stay on subsequent HRQoL (Davydow et al., 2008; Ringdal et al., 2009). Routine follow-up for restlessness during ICU treatment might help to identify patients in need of individualised care and, thus, increase the possibility of the nursing staff facilitating patients receiving the maximum benefit from treatment.

The finding that severe or unbearable pain experienced during an ICU stay has a negative effect on the post-operative HRQoL is to our knowledge new. This suggests that the management of pain should be one of the key areas of focus during the post-operative treatment period among cardiac surgery patients. However, our results must be regarded as preliminary and need to be confirmed in future.

6.8 Limitations of the study

The aim of the studies reported here was to evaluate the effect of different assumptions concerning the components of the quantification of QALYs in critical care setting. The results are based on empirical data gathered during an ordinary care process. The low response rate can be regarded as a limitation; but, since the objective of the studies was not to generalise the results to critical care setting (studies I, II and III), but instead to determine and illustrate the effect of calculation methods on the number of QALYs, the response rate is not a major cause of concern. The response rate in study IV, the objective of which differed from that for the other studies, was acceptable and the differences between respondents and non-respondents were relatively minor. Consequently, the results can be considered reliable.

The lack of a baseline HRQoL score is a major limitation related to the comparison of instruments' responsiveness to change (study I) and to the calculation of QALYs (study II). Obviously, the change in the HRQoL score would have been more prominent and the difference in the change score between instruments might have been more pronounced if it would have been assessed in a before–after design instead of evaluating HRQoL twice after treatment, i.e., at 6 and 12 months after ICU treatment. Despite this, the results here revealed differences between the instruments in several respects. In retrospect, proxy baseline assessments should have been performed in more than 100 patients, since the number of such evaluations was rather low in some of the diagnostic groups. As a consequence, our baseline assessment results should not be used as a standard for critical care patients, but rather as a theoretical example only. Regardless, the results of the baseline assessment reflect the differences between the instruments and show that baseline HRQoL scores may vary according to the diagnostic group of acutely ill patients.

The rule of rescue applies in critical care and, therefore, we do not know what would happen to patients in terms of the length and quality of life through "conventional" treatment alone. This is also difficult to establish, since it would be unethical to organise a trial where patients were randomised to receive ICU/HDU or conventional treatment in most cases. Consequently, studies evaluating the effectiveness of critical care must always be based on assumptions. To be able to compare the cost-effectiveness of treatments across a variety of medical specialties, similar assumptions should be used irrespective of whether care is provided on an emergency or elective basis. Because it is difficult to say which assumption set is most realistic in the critical care setting, there is a clear need for sensitivity analyses applying assumption sets when reporting the results of studies.

There was a slight selection bias in study I, since the inclusion criterion was survival time of at least 12 months and, as is known, 1-year mortality is elevated in the target patient group. This might have an impact on the magnitude of the observed ceiling effect using the EQ-5D, but not on the interpretation of differences between instruments. On the other hand, the slightly selected patient population—i.e., patients with quite a good survival rate—had scores consistent with health states WTD indicating that the EQ-5D can produce negative HRQoL scores even in fairly well-off patients. The selected data might explain the lack of very low HRQoL scores using the 15D.

In study II, hospital admission was assumed to have occurred on an emergency basis if the admission to both the hospital and ICU or HDU occurred on the same day. Consequently, the study population may also include some scheduled surgical patients who were admitted to the hospital on the day of the procedure, an conversely some acutely ill hospitalised patients may have been excluded from the data set. The comparison of the diagnoses observed in studies I and II, however, suggests that most of the patients were acutely ill in study II.

The follow-up time was restricted to a maximum of 12 months in studies dealing with HRQoL. Particularly in the study dealing with QALY calculations (study II), the restricted follow-up time in addition to other assumptions applied materialises as high cost per QALY ratios. However, the objective of this particular study was not to establish the cost utility of critical care, but to demonstrate the effects of

differences between calculation methods and instruments. A follow-up time of 12 months, instead of the 6 months used in study III, might have allowed for a better assessment of the permanence of the experienced restlessness and severe or unbearable pain. Due to the preliminary character of the study, this could not be anticipated beforehand. Additionally, the prevalence of experienced restlessness and severe or unbearable pain may have been higher if they had been assessed during their stay in the ward as well.

In study IV, the initial patient population included numerous patients; but, the high 30-day mortality rate reduced the patient population and analyses of narrow age groups were not possible for all diagnostic groups, especially, women. If the patient population had been large enough, the presented values in the tables concerning life expectancy among narrow age groups could have been even more useful for other studies within critical care settings.

6.9 Clinical implications

The findings here corroborate the notion that QALY is not a universal measure. This should be taken into account when evaluating the effectiveness of different forms of care. In addition, the results here expand the knowledge concerning the effect of methods used in QALY calculations and, thus, promote the critical evaluation of published cost-utility studies and the design of future cost-utility studies.

Reporting just the average number of QALYs gained may be insufficient, since in general there are patients for whom their HRQoL improves, remains unchanged or even deteriorates. As a consequence, the objective of utility studies should be, in addition to analysing the patient population as a group, to determine the reasons for the variation in treatment results between patients. It is important to establish which factors explain patient-specific variation. That is, are they explained by patient-related factors or are they explained by care process–related factors. The identification of such factors would likely improve patient selection and promote the development of the care process.

It is essential to understand that the HRQoL scores produced using different instruments are not interchangeable. In addition, the quite congruent average HRQoL scores produced using various instruments might conceal largely divergent distributions. Consequently, results from different HRQoL instruments cannot be combined and the distribution of HRQoL scores should be reported preferably in a graphical form as well.The annual relative survival method takes into account the mortality of the general population. Consequently, the excess mortality related to treatment can be identified without obtaining cause of death data. This is valuable with respect to elderly persons whose mortality from natural causes is high.

6.10 Future studies

There is a demand for future studies concerning the baseline HRQoL scores used in the critical care setting, the long-term mortality and HRQoL-related indicators. The patient-reported baseline HRQoL is impossible to obtain from all patients within critical care, necessitating a different solution to this particular problem. One possibility to resolve this issue might be to determine baseline values based on age group, sex and diagnosis and using these values across studies. However, the comparability between acutely ill and scheduled patients should be guaranteed.

Generally, funding for such studies is limited resulting in limited follow-up times and the need to resort to incomplete data sets. To resolve this particular problem, the compilation of a life table based on retrospective data might be useful for the determination of life expectancy for different studies. The life table should include diagnosis, sex, age group, the duration of excess mortality and extrapolated life expectancy.

There is some evidence that the follow-up HRQoL is predicted by factors other than those which predict morbidity and mortality. Since HRQoL is in itself an important objective of health care, it is essential to establish more conclusively in future studies which factors before and during treatment are important predictors of follow-up HRQoL.

Some HRQoL instruments—in particular, the EQ-5D—produce a negative HRQoL score for many patients, which imply that their health state is worse than being dead. Considering the credibility of such scores and of the entire instrument, it would be necessary to carry out a study where patients who have obtained a negative score would be asked directly whether they agree that their health state is worse than being dead and whether they would rather die than go on living in their present health state.

Different types of modelling—e.g., decision trees, Markov models and Monte Carlo simulations—are not yet commonly used to estimate cost utility in critical care settings. The data for modelling studies—i.e., information on outcomes, their probabilities, HRQoL scores, QALYs and costs associated with specific outcomes over time—are usually collected from diverse sources. It would be interesting to compare the results of such studies with those of prospective follow-up studies. Yet, the conclusions regarding the measurement of HRQoL and QALYs in this study also apply to modelling studies.

7. Conclusions

Studies based on empirical data demonstrated that QALY is not a universal measure. Instead, it is affected by how the factors to be taken into account in the calculation of QALYs are chosen and defined. Therefore, the calculation methods of QALYs should ideally be standardised. This may be difficult to achieve. At the very least, in each study using QALYs, the components used in their calculation should be clearly reported.

The methods and assumptions used in QALY calculations vary from study to study making comparisons between different studies difficult if not altogether impossible. When reporting the number of QALYs in a critical care setting, as a minimum the following elements should always be reported: how the baseline HRQoL was assessed, in which way recovery was assumed to take place, what calculation method was used (i.e., QALYs experienced or gained) and what measurement points including follow-up time and time horizon were used. When reporting the cost per QALY ratio, both the average or incremental as well as the costing methodology should be specified — that is, which resource items were included and how they were valued.

The method for assessing QALYs gained should be favoured over those methods which assess QALYs experienced, and the measurement points used should relate to the expected recovery of patients. The ranking of different health-care interventions in terms of their effectiveness calls for standardisation in the calculation of QALYs. The ranking of different interventions in terms of their cost utility (average cost-utility ratio) requires additional standardisation of the costing methodology. If societies define thresholds for acceptable incremental cost-utility ratios, they should be HRQoL instrument–specific given that different instruments, when used concurrently, produce different estimates for QALYs gained.

Factors affecting the follow-up HRQoL also influence the number of QALYs gained or experienced. Research to determine such factors should be carried out among different patient populations and environments. To improve the transparency and usefulness of HRQoL studies, the distribution of HRQoL scores and the proportion of patients who benefited from treatment as well as those who did not should be reported.

From the point of view of the QALY concept, negative HRQoL scores are problematic. The negative scores cause illogical outcomes and are difficult to interpret and act upon. In the field of health economics, consensus is needed in order to resolve these issues.

The annual RSR and the extrapolation of life expectancy are valuable methods in the estimation of life expectancy especially in patient populations with a high mortality rate and in ageing populations. Such methods increase the precision of QALY calculations.

Acknowledgements

The present study was carried out at the Hospital District of Helsinki and Uusimaa. I wish to thank everyone who made this work possible. Special thanks goes to the staff and patients whose participation made this work possible.

I'd like to extend my deepest gratitude to the following individuals:

Professor Emeritus Harri Sintonen, Ph.D. and Professor Risto P. Roine, M.D., Ph.D, the supervisors of this thesis. I have had the honour and privilege of working with these two authorities, who are both pioneers in the field of measuring Health Related Quality of Life and evaluating the effectiveness of health services in Finland. Their extensive expertise and mature perspectives on research and health economics have fundamentally impacted my work. In particular, I would like to thank Prof. Sintonen for accepting me as a doctoral student and Prof. Roine for persistently correcting my English. This work would not have been possible without their support and commitment.

Professor Tero Ala-Kokko M.D., Ph.D. and Docent Juha Laine Ph.D, the reviewers of this thesis. Prof. Ala-Kokko assessed this thesis from the perspective of intensive care medicine and Docent Laine evaluated it from the perspective of health economics. Their invaluable and constructive criticism and comments from their respective disciplines specifically improved the completeness of this work.

Professor Ville Pettilä M.D., Ph.D., Docent Raili Suojaranta-Ylinen M.D., Ph.D., Docent Pirjo Räsänen, Ph.D., Docent Antti Vento M.D., Ph.D., Adjunct Professor Irma-Leena Notkola, Ph.D. and Karri Seppä, Ph.D., the other co-authors of the original publications of this thesis. Each of these individuals contributed to the original articles and provided scientific vision, which improved upon the manuscripts immensely. Special gratitude is extended to Raili Suojaranta-Ylinen and Pirjo Räsänen from their encouraging and helpful collaboration and to Irma-Leena Notkola for her exceptional scientific and humanistic viewpoint.

the other members of the HUS QoL Study Group; Professor Olli-Pekka Ryynänen, M.D., Ph.D., Professor Marja Blom, Ph.D. and Pasi Aronen M.Sc. (Economics), who each engaged in rewarding and reasoned discussions concerning the topic of this thesis. My thanks also go to Ms. Heli Sarpila, who provided excellent technical assistance and Ms. Vanessa Fuller, M.A., who revised the language of this thesis.

all of my friends and colleagues who showed a genuine interest in my work and who provided support, friendship and joy throughout the process.

my husband's relatives from the summer villa in Kuortane, who afforded me an enjoyable and relaxing break from my typical routines.

my sister and her family, who supplied unfailing encouragement, support and many nice evenings spent together in town, at the summer villa or in Pärnu.

my children Marika, Minna and Joonas, my sons-in-law Mikko and Janne and my grandchildren Aada and Lukas. All of you are so significant to me that without you life would be miserable.

finally and most of all, my dear husband, Jukka, who has been my closest and best friend for over 30 years. Your support, patience, IT assistance and creative and relaxing humour make life much more comfortable.

This study was financially supported through grants from the District Hospital of Helsinki and Uusimaa and from the Yrjö Jahnsson Foundation.

Helsinki, April 2014

References

Agarwal S, Banerjee S, Tuzcu EM, Kapadia SR. Influence of age on revascularization related costs of hospitalization among patients of stable coronary artery disease. Am J Cardiol 2010;105:1549-54.

Alanne S. Subjektiivisesti merkitsevä muutos terveyteen liittyvässä elämänlaadussa. Itä-Suomen yliopisto, Sosiaali-ja terveysjohtamisen laitos, Pro gradu 2011.

Al-Ruzzeh S, Epstein D, George S, Bustami M, Wray J, Ilsley C, Sculpher M, Amrani M. Economic evaluation of coronary artery bypass grafting surgery with and without cardiopulmonary bypass: cost-effectiveness and quality-adjusted life years in a randomized controlled trial. Artif Organs 2008;32:891-7.

Angus DC, Carlet J. Surviving intensive care: a report from the 2002 Brussels Roundtable. Intensive Care Med 2003;29:368–77.

Almerud S, Alapack RJ, Fridlund B, Ekebergh M. Of vigilance and invisibility- being a patient in technologically intense environments. Nurs Crit Care 2007;12:151-8.

Azzopardi S, Lee G. Health-related quality of life 2 years after coronary artery bypass graft surgery. J Cardiovasc Nurs 2009;24:232-40.

Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and time trade-off. Qual Life Res 1999;8:303-10.

Badia X, Diaz-Prieto A, Gorriz MT, Herdman M, Torrado H, Farrero E, Cavanilles JM. Using the EuroQol-5D to measure changes in quality of life 12 months after discharge from an intensive care unit. Intensive Care Med 2001;27:1901-7.

Bhamidipati CM, LaPar DJ, Fonner E Jr, Kern JA, Kron IL, Ailawadi G. Outcomes and cost of cardiac surgery in octogenarians is related to type of operation: a multi-institutional analysis. Ann Thorac Surg 2011;91:499-505.

Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. Health Technol Assess 1999;3:55-69.

Bland JM, Altman DG.Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

Bleichrodt H, Johannesson M. The Validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Med Decis Making 1996;17:21-32.

Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. Value Health 2006;9:213-8.

Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. J Health Serv Res Policy 1999;4:174-84.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.

Brooks R, the EuroQol Group. EuroQol: the current state of play. Health Poli 1996;37:53-72.

Browne JP, van der Meulen JH, Lewsey MJ, Lamping DL, Black N. Mathematical coupling may account for association between baseline severity and minimally important difference values. J Clin Epidemiol 2010;63:865-74.

Chatterji S, Ustün BL, Sadana R, Salomon JA, Mathers CD, Murray CJL. The conceptual basis for measuring and reporting on health. Global Programme on Evidence for Health Policy Discussion Paper No 45, WHO 2002.

Cheung AM, Tansey CM, Tomlinson G, Diaz-Granados N, Matté A, Barr A, Mehta S, Mazer CD, Guest CB, Stewart TE, Al-Saidi F, Cooper AB, Cook D, Slutsky AS, Herridge MS. Two-year outcomes, health care use, and costs of survivors of acute respiratory distress syndrome. Am J Respir Crit Care Med 2006;174:538-44.

Coast J. Reprocessing data to form QALYs. BMJ 1992;305:87-90.

Cohen BJ. Discounting in cost-utility analysis of healthcare interventions: reassessing current practice. Pharmacoecon 2003;21:75-87.

Cohen DJ, Lavelle TA, Van Hout B, Li H, Lei Y, Robertus K, Pinto D, Magnuson EA, Mcgarry TF, Lucas SK, Horwitz PA, Henry CA, Serruys PW, Mohr FW, Kappetein AP. Economic outcomes of percutaneous coronary intervention with drug-eluting stents versus bypass surgery for patients with left main or three-vessel coronary artery disease: one-year results from the SYNTAX trial. Catheter Cardio Inte 2012;79:198-209.

Cole RP, Shakespeare V, Shakespeare P, Hobby JA. Measuring outcome in low-priority plastic surgery patients using quality of life indices. Br J Plast Surg 1994;47:117-21.

Coons SJ, Rao S, Keininge DL, Hay RDA. Comparative review of generic quality-of-life instruments. Pharmacoecon 2000;17:13-35.

Cuthbertson BH, Roughton S, Jenkinson D, Maclennan G, Vale L. Quality of life in the five years after intensive care: a cohort study. Crit Care 2010;14:R6.

Dasta JF, McLaughlin TP, Mody SH, Piech CT. Daily costs of an intensive care unit: the contribution of mechanical ventilation. Crit Care Med 2005;33:1266-71.

Davydow DS, Gifford JM, Desai SV, Bienvenu OJ, Needham DM. Depression in general intensive care unit survivors: a systematic review. Intensive Care Med 2009;35:796-809.

Deja M, Denke C, Weber-Carstens S, Schröder J, Pille CE, Hokema F, Falke KJ, Kaisers U. Social support during intensive care unit stay might improve mental impairment and consequently health-related quality of life in survivors of severe acute respiratory distress syndrome. Crit Care 2006;10:R147.

Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996;5:141-54.

Dolan, P. Modeling valuations for EuroQol health states. Medical Care 1997;35:1095-108.

Dolan P. Utilitarianism and the measurement and aggregation of quality-adjusted life years. Health Care Anal 2001;9:65-76.

Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximization and people's preferences: a methodological review of the literature. Health Econ 2005;14:197-208.

Eefting F, Nathoe H, van Dijk D, Jansen E, Lahpor J, Stella P, Suyker W, Diephuis J, Suryapranata H, Ernst S, Borst C, Buskens E, Grobbee D, de Jaegere P. Randomized comparison between stenting and off-pump bypass surgery in patients referred for angioplasty. Circulation 2003;108:2870-6.

Elliott D, Mudaliar Y, Kim C. Examining discharge outcomes and health status of critically ill patients: some practical considerations. Intensive Crit Care Nurs 2004;20:366-77.

Ely EW, Truman B, Shintani A, Thomason JW, Wheeler AP, Gordon S, et al. Monitoring sedation status over time in ICU patients: reliability and validity of the Richmond Agitation-Sedation Scale (RASS). JAMA 2003;289:2983-91.

EQ-5D concepts and methods: A developmental history. Kind P, Brooks R, Rabin R (Eds.) Springer, Netherlands 2005:19-27.

The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Poli 1990;16:199-208.

EuroQol Group. EQ-5D-5L users Guide, Version 1.0.
http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L.pdf

Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUP) system for assessing health-related quality of life in clinical studies. Ann Med 2001:33;375-84.

Färkkilä N, Torvinen S, Roine RP, Sintonen H, Hänninen J, Taari K, Saarto T. Health-related quality of life among breast, prostate, and colorectal cancer patients with end-stage disease. Qual Life Res. 2013;23:1387-94

Gaudino M, Girola F, Piscitelli M, Martinelli L, Anselmi A, Della Vella C, Schiavello R, Possati G. Long-term survival and quality of life of patients with prolonged postoperative intensive care unit stay: unmasking an apparent success. J Thorac Cardiovasc Surg 2007;134:465-9.

Gelsomino S, Lorusso R, Livi U, Masullo G, Lucà F, Maessen J, Gensini GF. Cost and cost-effectiveness of cardiac surgery in elderly patients. J Thorac Cardiovasc Surg 2011;142:1062-73.

Gersbach P, Tevaearai H, Revelly JP, Bize P, Chioléro R, von Segesser LK. Are there accurate predictors of long-term vital and functional outcomes in cardiac surgical patients requiring prolonged intensive care? Eur J Cardiothorac Surg 2006;29:466-72.

Gjeilo KH, Wahba A, Klepstad P, Lydersen S, Stenseth R. Recovery patterns and health-related quality of life in older patients undergoing cardiac surgery-a prospective study. Eur J Cardiovasc Nurs 2012;11:322-30.

Grabbe PFM, Essink-Bot M-L, Bonsel GJ. The comparability and reliability of five health-state valuation methods. Soc Si Med 1997;45:1641-52.

Grady KL, Lee R, Subačius H, Malaisrie SC, McGee EC Jr, Kruse J, Goldberger JJ, McCarthy PM. Improvements in health-related quality of life before and after isolated cardiac operations. Ann Thorac Surg 2011;91:777-83.

Graf J, Wagner J, Graf C, Koch KC, Janssens U. Five-year survival, quality of life, and individual costs of 303 consecutive medical intensive care patients-a cost-utility analysis. Crit Care Med 2005;33:547-55.

Graf J, Mühlhoff C, Doig GS, Reinartz S, Bode K, Dujardin R, Koch KC, Roeb E, Janssens U. Health care costs, long-term survival, and quality of life following intensive care unit admission after cardiac arrest. Crit Care 2008;12:R92.

Green C, Brazier J, Deverill M. Valuing Health-Related Quality of Life. A review of health state valuation techniques. Pharmacoeco 2000;17:151-65.

Greene WH. LIMDEP Version 7.0: User's manual, revised version. Econometric Software, Inc.: New York, 1998

Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, Badi X, Busschbach J, Buxton M, Dolan P, Kind P, Krabbe P, Ohinmaa A, Parkin D, Rose M, Sintonen H, Tsuchiya A, de Charro F. A single European currency for EQ-5D health states. Results from a six-country study. Eur J Health Econom 2003;4:222–31.

Guide to the Methods of Technology Appraisal. National Institute for Clinical Exellence. June 2008. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

Halpern NA, Pastores SM. Critical care medicine in the United States 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010;38:65-71.

Hamel MB, Phillips RS, Davis RB, Teno J, Connors AF, Desbiens N, Lynn J, Dawson NV, Fulkerson W, Tsevat J. Outcomes and cost-effectiveness of ventilator support and aggressive care for patients with acute respiratory failure due to pneumonia or acute respiratory distress syndrome. Am J Med 2000;109:614-20.

Harris A, Cooper BA, Li JJ, Bulfone L, Branley P, Collins JF, Craig JC, Fraenkel MB, Johnson DW, Kesselhut J, Luxton G, Pilmore A, Rosevear M, Tiller DJ, Pollock CA, Harris DC. Cost-effectiveness of initiating dialysis early: a randomized controlled trial. Am J Kidney Dis 2011;57:707-15.

Hawkes AL, Mortensen OS. Up to one third of individual cardiac patients have a decline in quality of life post-intervention. Scand Cardiovasc J 2006;40:214-8.

Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual Life Res1999;8:209-24.

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.

Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med 2001;33:350-7.

Van den Heede K, Lesaffre E, Diya L, Vleugels A, Clarke SP, Aiken LH, Sermeus W. The relationship between inpatient cardiac surgery mortality and nurse numbers and educational level: analysis of administrative data. Int J Nurs Stud 2009;46:796-803.

Heijink R, van Baal P, Oppe M, Koolman X, Westert G. Decomposing cross-country differences in quality adjusted life expectancy: the impact of value sets. Popul Health Metr 2011;9:17.

Hein OV, Birnbaum J, Wernecke K, England M, Konertz W, Spies C. Prolonged intensive care unit stay in cardiac surgery: risk factors and long-term-survival. Ann Thorac Surg 2006;81:880-5.

Hofhuis JG_1, Spronk PE, van Stel HF, Schrijvers GJ, Rommes JH, Bakker J. The impact of critical illness on perceived health-related quality of life during ICU treatment, hospital stay, and after hospital discharge: a long-term follow-up study. Chest 2008;133:377-85.

Hofhuis JG_2, Spronk PE, van Stel HF, Schrijvers AJ, Rommes JH, Bakker J. The impact of severe sepsis on health-related quality of life: a long-term follow-up study. Anesth Analg 2008;107:1957-64.

Honkalampi T, Sintonen H. Do the 15D scores and time trade-off (TTO) values of hospital patients' own health agree? Int J Technol Assess Health Care 2010;26:117-23.

Intensium Benchmarking database, Tieto Healthcare & Welfare. http://www.intensium.com/web/english/

Iribarne A, Easterwood R, Russo MJ, Chan EY, Smith CR, Argenziano M. Comparative effectiveness of minimally invasive versus traditional sternotomy mitral valve surgery in elderly patients. Thorac Cardiovasc Surg 2012;143:S86-90.

Jokinen JJ, Hippeläinen MJ, Turpeinen AK, Pitkänen O, Hartikainen JE. Health-related quality of life after coronary artery bypass grafting: a review of randomized controlled trials. J Card Surg 2010;25:309-17.

Kaarlola A, Pettilä V, Kekki P. Quality of life six years after intensive care. Intensive Care Med 2003;29:1294-9.

Kaarlola A, Tallgren M, Pettilä V. Long-term survival, quality of life, and quality-adjusted life-yearsamong critically ill elderly patients. Crit Care Med 2006;34:2120-6.

Kantola T, Mäklin S, Koivusalo AM, Räsänen P, Rissanen A, Roine R, Sintonen H, Höckerstedt K, Isoniemi H. Cost-utility of molecular adsorbent recirculating system treatment in acute liver failure. World J Gastroenterol 2010;16:2227-34.

Karlsson S, Ruokonen E, Varpula T, Ala-Kokko TI, Pettilä V. Long term-outcome and quality-adjusted life years after serve sepsis. Crit Care Med 2009;34:1268-74.

Kattainen E. A longitudinal research of the coronary artery bypass grafting (CABG) and percutaneous transluminal coronary artery angioplasty (PTCA) patients' health-related quality of life. Kuopio University Publications E. Social Sciences 114. Kuopio 2004.

Kattainen E, Sintonen H, Kettunen R, Meriläinen P. Health-related quality of life of coronary artery bypass grafting and percutaneous transluminal coronary artery angioplasty patients: 1-year follow-up. Int J Technol Assess Health Care 2005;21(2):172-179.

Khouli H, Astua A, Dombrowski W, Ahmad F, Homel P, Shapiro J, Singh J, Nallamothu R, Mahbub H, Eden E, Delfiner J. Changes in health-related quality of life and factors predicting long-term outcomes in older adults admitted to intensive care units. Crit Care Med 2011;39:731-7.

King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171-84.

Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. Value Health 2009;12:1151-7.

Kontodimopoulos N, Pappa E, Chadjiapostolou Z, Arvanitaki E, Papadopoulos AA, Niakas D. Comparing the sensitivity of EQ-5D, SF-6D and 15D utilities to the specific effect of diabetic complications. Eur J Health Econ 2012;13:111-20.

Kopec JA, Willison KD. A comparative rewiev of four preference-weighted measures of health-related quality of life. J Clin Epidemiol 2003;56:317-25.

Kvale R, Flaatten H. Changes in health-related quality of life from 6 months to 2 years after discharge from intensive care. Health Qual Life Outcomes 2003;24: doi:10.1186/1477-7525-1-2.

Kvam AK, Fayers PM, Wisloff F. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. Eur J Haematol 2011;87:330–7.

Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.

Lillegraven S, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. Ann Rheum Dis 2012;69:1762-7.

Linde L, Sorensen J, Ostergaard M, Horslev-petersen K, Lund-Hetland M. Health-related quality of life: validity, reliability and responsiveness of SF-36, 15D, EQ-5D, RAQoL and HAQ in patients with rheumatoid arthritis. J Rheumatol 2008;35:1528-37.

Linko R, Suojaranta-Ylinen R, Karlsson S, Ruokonen E, Varpula T, Pettilä V. One-year mortality, quality of life and predicted life-time cost-utility in critically ill patients with acute respiratory failure. Crit Care 2010;14:R60

Loponen P, Luther M, Wistbacka J-O, Nissinen J, Sintonen H, Huhtala H, Tarkka MR. Postoperative delirium and health related quality of life after coronary artery bypass grafting. Scand Cardiovasc J 2008;42: 337-44.

Loponen P, Luther M, Korpilahti K, Wistbacka JO, Huhtala H, Laurikka J, Tarkka MR. HRQoL after coronary artery bypass grafting and percutaneous coronary intervention for stable angina. Scand Cardiovasc J 2009;43:94-9.

Mahoney EM, Greenberg D, Lavelle TA, Natarajan A, Berezin R, Ishak KJ, Caro JJ, Yadav JS, Gray WA, Wholey MH, Cohen DJ. Costs and cost-effectiveness of carotid stenting versus endarterectomy for patients at increased surgical risk: results from the SAPPHIRE trial. Catheter Cardio Inte 2011;77:463-72.

Malmivaara K, Kivisaari R, Hernesniemi J, Siironen J. Cost-effectiveness of decompressive craniectomy in traumatic brain injuries. Eur J Neurol 2011;18:656-62.

Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005;14:487-96.

Markou AL, de Jager MJ, Noyez L. The impact of coronary artery disease on the quality of life of patients undergoing aortic valve replacement. Interact Cardiovasc Thorac Surg 2011;13:128-32.

Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, Esdaile JM, Anis AH. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med 2005;60:1571-82.

Marra CA, Marion SA, Guh DP, Najafzadeh M, Wolfe F, Esdaile JM, Clarke AE, Gignac MA, Anis AH. Not all "quality-adjusted life years" are equal. J Clin Epidemiol 2007;60:616-24.

Mayer SA, Copeland D, Bernardini GL, Boden-Albala B, Lennihan L, Kossoff S, Sacco RL. Cost and outcome of mechanical ventilation for life-threatening stroke. Stroke 2000;31:2346-53.

Mayr VD, Dünser MW, Greil V, Jochberger S, Luckner G, Ulmer H, Friesenecker BE, Takala J, Hasibeder WR. Causes of death and determinants of outcome in critically ill patients. Crit Care 2006;10:R154.

Merlani P, Chenaud C, Mariotti N, Ricou B. Long-term outcome of elderly patients requiring intensive care admission for abdominal pathologies: survival and quality of life. Acta Anaesthesiol Scand 2007;51:530-7.

Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;17:485-95.

Moran JL, Peisach AR, Solomon PJ, Martin J. Cost calculation and prediction in adult intensive care: a ground-up utilization study. Anaesth Intensive Care 2004;32:787-97.

National Centre for Priority Setting in Health Care 2008. Resolving health care's difficult choices. Survey of priority setting in Sweden and an analysis of principles and guidelines on priorities in health care. Prioriterings Centrum 2008;2.

Niskanen M, Kari A, Halonen P. Five-years survival after intensive care- Comparison of 12,180 patients with the general population. Crit Care Med 1996;24:1962-7

Niskanen M, Reinikainen M, Pettilä V. Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. Intensive Care Med. 2009;35:1060-7.

Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health. 2009;12 Suppl 1:S10-5. doi: 10.1111/j.1524-4733.2009.00516.x.

Orwelius L, Nordlund A, Nordlund P, Simonsson E, Bäckman C, Samuelsson A, Sjöberg F. Pre-existing disease: the most important factor for health related quality of life long-term after critical illness: a prospective, longitudinal, multicentre trial. Crit Care 2010;14:R67.

Pappalardo F, Franco A, Landoni G, Cardano P, Zangrillo A, Alfieri O. Long-term outcome and quality of life of patients requiring prolonged mechanical ventilation after cardiac surgery. Eur J Cardiothorac Surg 2004;25:548-52.

Peek GJ, Elbourne D, Mugford M, Tiruvoipati R, Wilson A, Allen E, Clemens F, Firmin R, Hardy P, Hibbert C, Jones N, Killer H, Thalanany M, Truesdale A. Randomised controlled trial and parallel economic evaluation of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR). Health Technol Assess 2010;14:1-46.

Perez DJ, McGee R, Campbell AV, Christensen EA, Williams S. A comparison of time trade-off and quality of life measures in patients with advanced cancer. Qual Life Res. 1997;6:133-8.

Pesonen A, Suojaranta-Ylinen R, Tarkkila P, Rosenberg PH. Applicability of tools to assess pain in elderly patients after cardiac surgery. Acta Anaesthesiol Scand 2008;52:267-73.

Pharmaceuticals Pricing Board 2011.
http://www.stm.fi/c/document_library/get_file?folderId=71837&name=DLFE-15658.pdf

Pickard AS, Neary M.P, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007;5:70.

Pindyck RS, Rubinfeld DL. Econometric models and economic forecasts. McGraw-Hill, New York 1991, s. 240.

Prieto L, Sacristán JA. Problems and solutions in calculating quality-adjusted life years (QALYs). Health Qual Life Outcomes 2003;19:doi: 10.1186/1477-7525-1-80.

Pätilä T, Kukkonen S, Vento A, Pettilä V, Suojaranta-Ylinen R. Relation of the sequential organ failure assessment score to morbidity and mortality after cardiac surgery. Ann Thorac Surg 2006;82:2072-8.

Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337-43.

Ranta E, Rita H, Kouki J. Biometria tilastotiedettä ekologeille. Yliopistopaino, Helsinki 2005:206-210,136-43.

Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgements. BMJ 2004;329:224-7.

Reis M, Dinis MD, de Rijk A, Schaufeli W. Simplified Therapeutic Intervention Scoring System: The TISS-28 items-Results from a multicenter study. Crit Care Med 1996;24:64-73.

Revick D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important difference for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9.

Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ 2004;13:1203-10.

Richardson J, Lezzi A, Maxwell A. Cross-national comparison of twelve quality of life instruments. Monash University, Center for Health Economics. Research Paper 2012;76.

Rimachi R, Vincent JL, Brimioulle S. Survival and quality of life after prolonged intensive care unit stay. Anaesth Intensive Care 2007;35:62-7.

Ringdal M, Plos K, Lundberg D, Johansson L, Bergbom I. Outcome after injury: memories, health-related quality of life, anxiety, and symptoms of depression after intensive care. J Trauma 2009;66:1226-33.

Robinson JC. Variation in hospital costs, payments, and profitabilty for cardiac valve replacement surgery. Health Serv Res 2011;46:1928-45.

Rodriguez JM, Paz S, Lizan L, Gonzalez P. The use of quality-adjusted life-years in the economic evaluation of health technologies in Spain: a review of the 1990-2009 literature. Value Health 2011;14:458-64.

Rotondi AJ, Chelluri L, Sirio C, Mendelsohn A, Schulz R, Belle S, Im K, Donahoe M, Pinsky MR. Patients' recollections of stressful experiences while receiving prolonged mechanical ventilation in an intensive care unit. Crit Care Med 2002;30:746-52.

Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynänen OP, Roine R. 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;22:235-41.

Saarni SI, Härkänen T, Sintonen H, Suvisaari J, Koskinen S, Aromaa A, Lönnqvist J. The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. Qual Life Res 2006;15:1403-14.

Saarni SI, Viertiö S, Perälä J, Koskinen S, Lönnqvist J, Suvisaari J. Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. Br J Psychiatry 2010;197:386-94.

Sarna S (a). Kliinisen biostatiikan kurssi, 2011. http://www.kttl.helsinki.fi/sarna/osa1.pdf

Sarna S (b). Kliinisen biostatistiikan jatkokurssi, 2011. http://www.kttl.helsinki.fi/sarna/osa2.pdf

Saxena A, Poh CL, Dinh DT, Reid CM, Smith JA, Shardey GC, Newcomb AE. Early and late outcomes after isolated aortic valve replacement in octogenarians: an Australasian Society of Cardiac and Thoracic Surgeons Cardiac Surgery Database Study. Eur J Cardiothorac Surg 2012;41:63-8.

Schelling G, Richter M, Roozendaal B, Rothenhäusler HB, Krauseneck T, Stoll C, Nollert G, Schmidt M, Kapfhammer HP. Exposure to high stress in the intensive care unit may have negative effects on health-related quality-of-life outcomes after cardiac surgery. Crit Care Med 2003;31:1971-80.

Schenk P, Warszawska J, Fuhrmann V, König F, Madl C, Ratheiser K. Health-related quality of life of long-term survivors of intensive care: changes after intensive care treatment : experience of an Austrian intensive care unit. Wien Klin Wochenschr 2012;124:624-32

Schwappach DL, Boluarte TA. HEE-GER: a systematic review of German economic evaluations of health care published 1990-2004. BMC Health Serv Res 2007;12: doi:10.1186/1472-6963-7-7.

Scuffham PA, Whitty JA, Mitchell A, Viney R. The use of QALY weights for QALY calculations: a review of industry submissions requesting listing on the Australian pharmaceutical benefits scheme 2002-4. Pharmacoecon 2008;26:297-310.

Seferian EG, Afessa B. Demographic and clinical variation of adult intensive care unit utilization from geographically defined population. Crit Care Med 2006;34:2113-9.

Seppä K, Hakulinen T. Mean and median survival times of cancer patients should be corrected for informative censoring. J Clin Epidemiol 2009;62:1095-102

Shaw JW, Pickard AS, Yu S, Chen S, Iannacchione VG, Johnson JA, Coons SJ. A median model for predicting United States population-based EQ-5D health state preferences. Value Health 2010;13:278-88.

Sintonen H. The 15-D Measure of health-related quality of life: Reliability, validity and sensitivity of its health state descriptive system. Melbourne, Centre for Health Program Evaluation. Working Paper 41, 1994.

Sintonen H. The 15-D Measure of health-related quality of life: II Feasibility, reliability and validity of its valuation system. Melbourne, Centre for Health Program Evaluation. Working Paper 42, 1995.

Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;33:328-36.

Sintonen H, Weijnen T, Nieuwenhuizen M, Oppe S et al. Comparison of EQ-5D VAS valuations: analysis of background variables. In Brooks, R, Rabin R, de Charro F (eds) The measurement and valuation of health status using the EQ-5D: A European perspective. Evidence from the EuroQol BIOMED Research programme. Kluwer Academic Publishers, Dordrecht 2003, 81-101.

Statistics Finland, database

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.

Stavem K, Bjørnaes H, Lossius MI. Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. Epilepsy Res 2001;44:179-89.

Stavem K, Frøland SS, Hellum KB. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Qual Life Res 2005;14:971-80.

Stricker KH, Cavegn R, Takala J, Rothen HU. Does ICU length of stay influence quality of life? Acta Anaesthesiol Scand 2005;49:975-83.

Stricker KH, Sailer S, Uehlinger DE, Rothen HU, Zuercher Zenklusen RM, Frick S. Quality of life 9 years after an intensive care unit stay: a long-term outcome study. J Crit Care 2011;26:379-87.

Sultan S, Hynes N. Clinical efficacy and cost per quality-adjusted life years of pararenal endovascular aortic aneurysm repair compared with open surgical repair. J Endovasc Ther 2011;18:181-96.

Suojaranta-Ylinen RT, Kuitunen AH, Kukkonen SI, Vento AE, Salminen US. Risk evaluation of cardiac surgery in octogenarians. J Cardiothorac Vasc Anesth 2006;20:526-30.

Sznajder M, Aegerter P, Launois R, Merliere Y, Guidet B, CubRea. A cost-effectiveness analysis of stays in intensive care units. Intensive Care Med 2001;27:146-53.

Talmor D, Greenberg D, Howell MD, Lisbon A, Novack V, Shapiro N. The costs and cost-effectiveness of an integrated septicaemia treatment protocol. Crit Care Med 2008;36:1168-74.

Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. Guidelines for intensive care unit admission, discharge, and triage. Crit Care Med 1999;27:633-8.

Timmers TK, Verhofstad MH, Moons KG, van Beeck EF, Leenen LP. Long-term quality of life after surgical intensive care admission. Arch Surg 2011;146:412-8.

Torvinen S, Färkkilä N, Sintonen H, Saarto T, Roine RP, Taari K. Health-related quality of life in prostate cancer. Acta Oncol. 2013 Aug;52:1094-101

Trouillet JL, Luyt CE, Guiguet M, Ouattara A, Vaissier E, Makri R, Nieszkowska A, Leprince P, Pavie A, Chastre J, Combes A. Early percutaneous tracheotomy versus prolonged intubation of mechanically ventilated patients after cardiac surgery: a randomized trial. Ann Intern Med 2011;154:373-83.

Vaara ST, Pettilä V, Reinikainen M, Kaukonen KM, Finnish Intensive Care Consortium. Population-based incidence, mortality and quality of life in critically ill patients treated with renal replacement therapy: a nationwide retrospective cohort study in Finnish intensive care units. Crit Care 2012;16:R13.

Valentin A, Ferdinande P. ESICM Working Group on Quality Improvement. Recommendations on basic requirements for intensive care units: structural and organizational aspects. Intensive Care Med 2011;37:1575-87.

Vest MT, Murphy TE, Araujo KL, Pisani MA. Disability in activities of daily living, depression, and quality of life among older medical ICU survivors: a prospective cohort study. Health Qual Life Outcomes 2011;9: doi: 10.1186/1477-7525-9-9.

Wachter RM, Luce JM, Safrin S, Berrios DC, Charlebois E, Scitovsky AA. Cost and outcome of intensive care for patients with AIDS, Pneumocystis carinii pneumonia, and severe respiratory failure. JAMA 1995;273:230-5.

Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.

Wang K, Zahang B, Li C, Wang C. Qualitative analysis of patients' intensive care experience during mechanical ventilation. J Clin Nurs 2008;18:183-90.

Ware JE Jr. SF-36 Health Survey Update. Spine 2000;25:3130–9.

Wehler M, Geise A, Hadzionerovic D, Aljukic E, Reulbach U, Hahn EG, Strauss R. Health-related quality of life of patients with multiple organ dysfunction: individual changes and comparison with normative population. Crit Care Med 2003;31:1094-101.

Weintraub WS, Mahoney EM, Zhang Z, Chu H, Hutton J, Buxton M, Booth J, Nugara F, Stables RH, Dooley P, Collinson J, Stuteville M, Delahunty N, Wright A, Flather MD, De Cock E. One year comparison of costs of coronary surgery versus percutaneous coronary intervention in the stent or surgery trial. Heart 2004;90:782-8.

Welsby IJ, Bennett-Guerrero E, Atwell D, White WD, Newman MF, Smith PK, Mythen MG. The association of complication type with mortality and prolonged stay after cardiac surgery with cardiopulmonary bypass. Anesth Analg 2002;94:1072-8.

Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull 2010;96:5-21.

World Health Organization. The first ten years of the World Health Organization. WHO, Geneva 1958, p. 459.

Wu Y, Jin R, Gao G, Grunkemeier GL, Starr A. Cost-effectiveness of aortic valve replacement in the elderly: an introductory study. J Thorac Cardiovasc Surg 2007;133:608-13.

Wyrwich KW. Minimal Important Difference thresholds and the standard error of measurement: is there a connection? J Biopharm Statist 2004;14:97–110.

Vaara ST, Pettilä V, Reinikainen M, Kaukonen KM; the Finnish Intensive Care Consortium.Population-based incidence, mortality and quality of life in critically ill patients treated with renal replacement therapy: a nationwide retrospective cohort study in Finnish intensive care units. Crit Care 2012;16:R13.

Zannad F, Mebazaa A, Juillière Y, Cohen-Solal A, Guize L, Alla F, Rougé P, Blin P, Barlet MH, Paolozzi L, Vincent C, Desnos M, Samii K; EFICA Investigators. Clinical profile, contemporary management and one-year mortality in patients with severe acute heart failure syndromes: The EFICA study. Eur J Heart Fail 2006;8:697-705.

Yaghoubi M, Aghayan HR, Arjmand B, Emami-Razavi SH. Cost-effectiveness of homograft heart valve replacement surgery: an introductory study. Cell Tissue Bank 2011;12:153-8.