Quality of Life depending on assessment method:

The EQ-5D Descriptive System versus Direct Time-Trade-Off Grading

Carl Willers*
Bachelor's Thesis, Stockholm School of Economics, Department of Economics

Abstract

When deciding upon resource allocation within the health care sector, implications on patients’ health-related Quality of Life (QoL) are often of central concern. Different instruments with the purpose of assessing perceived QoL are in use, amongst other the EQ-5D descriptive system. It was based on the method of Time-Trade-Off (TTO) grading by healthy individuals regarding hypothetical, described states of impaired health. However, it has been of interest to investigate whether the Quality-of-Life values of EQ-5D corresponds to those of directly elicited TTO, especially since previous empirical research has pointed out discrepancies. The aim of this thesis was to compare outcomes from the two instruments and investigate if they differ, and to put the instrument outcomes in a practical context through simulating a cost-utility analysis. Therefore, results from both instruments regarding one cohort of patients were compared and tested for statistically significant differences. Analyses showed that such were of apparent existence, mean outcomes from direct TTO being consistently higher than mean outcomes from EQ-5D. To put this difference in a practical context, a previously published cost-effectiveness model was used to compute incremental cost-effectiveness ratios (ICERs) with each QoL-assessing method respectively, evaluating a hypothetical drug compared to no treatment for osteoporosis patients. Due to the difference in QoL outcomes from the two instruments, the ICERs were also differing. Using QoL values from the direct TTO method implied the treatment to be regarded as less cost-effective compared to when QoL estimates based on the EQ-5D descriptive system were used.

Keywords: Quality of Life, Time-Trade-Off grading, EQ-5D, Comparative Study, Cost-Utility Analysis

Tutor: Magnus Johannesson
*20952@live.hhs.se
**List of Figures**

Figure 1. Overview of the preference-based technique used for Time-Trade-Off grading.....7
Figure 2. Timeline for data collection ........................................................................................................12
Figure 3. A Markov model with possible health states and transitions visualized. ..............18
Figure 4. Flow chart showing exclusion of patients from collected sample .................................25

**List of Tables**

Table 1. Baseline characteristics for subset of patients .................................................................21
Table 2. Overview of mean estimated Quality of Life .................................................................22
Table 3. Value set overview ..........................................................................................................24
Table 4. Examples of health states and their quality weights depending on method ...............24
Table 5. Baseline characteristics for subset of patients .................................................................26
Table 6. Utility multipliers for the Markov states in the model ....................................................26
Table 7. Incremental QALY gains and ICERs ...............................................................................26
Glossary

EQ-5D™ European Quality of life in 5 Dimensions, a descriptive questionnaire for grading perceived quality of life. Trade mark of the EuroQol Group which also maintains the instrument

HRQoL Health-Related Quality of Life
In this thesis, no distinction is made between HRQoL and QoL

Inpatient Treated at hospital

ICER Incremental Cost-Effectiveness Ratio

NICE National Institute for Clinical Excellence (United Kingdom)

Outpatient Treated at home or through visits at health care establishments

TLV The Dental and Pharmaceutical Benefits Agency in Sweden

TTO Time-Trade-Off

QALY Quality-Adjusted Life-Year, an aggregated measure referring to quantity of life as well as perceived quality of life

QoL Quality of Life

WTP Willingness To Pay, may apply to therapies or drugs regarding how much authorities are willing to spend on a specific treatment, often on per-QALY basis
1 Introduction

Financing of health care is highly prioritized by European governments, as public health expenditures make up a large part of overall resources allocated to health care [18]. The concept of value for money is of central concern for purchasers of health care in this matter, since the ability to improve people's health and life-quality is limited by scarce resources. Thus, evidence is needed – proving that a medical intervention is actually enhancing the patients treated. Economic evaluation of such interventions is therefore in demand [18].

A common method for performing economic evaluations of medical interventions is cost-effectiveness analysis (CEA). The CEA enables comparison of costs associated with different treatment options to the medical improvement associated with the treatments evaluated [12,18]. For such a comparison to be valid, there is a need for rational assessment methods regarding the medical improvement [7]. Since several treatments do not enhance survival although enhancing perceived Quality of Life (QoL), a measure of changes in QoL is essential [24].

QoL has been defined as a subjective matter that is dependent on objective human needs [5]. In this study, QoL refers to the state of physical, mental and emotional wellbeing, as related to the World Health Organization’s definition of health [34]. No distinction is made between QoL and Health-Related Quality of Life (HRQoL). There are both subjective and objective approaches to try assessing individuals’ perceived QoL [5]. Several instruments have been developed with this particular purpose – disease-specific as well as those applicable to any disease or condition (i.e. generic) [4,19].

Direct Time-Trade-Off (TTO) grading is one generic method for evaluating health states.. It is a choice-based technique, reflecting the respondent’s preferences regarding life-years in full health versus life-years in impaired health. The technique of direct TTO has furthermore been used as the basis for development of additional instruments for QoL assessment.

European Quality of life in five dimensions (EQ-5D descriptive system, EQ-5D) is an instrument that is based on direct TTO and may be referred to as indirect TTO since one’s grading of

1
EQ-5D is translated into a utility value corresponding to that of TTO. The EQ-5D consists of a questionnaire with questions regarding five dimensions of daily life, each dimension to be graded on a three-level scale. Based on a previously performed study [7], every possible health state determined by this questionnaire can be assigned a health-related utility value [7,10]. The assignment of utility values to each health state was enabled by letting a study sample perform TTO evaluations of health states that were described in terms of the EQ-5D descriptive system; outcomes from the two methods were hence possible to link. Such a link may also be referred to as an EQ-5D value set or algorithm, assigning a utility value associated with each level of EQ-5D’s dimensions. However, the study sample during EQ-5D elaboration did not consist of actual patients, but rather a sample of healthy individuals grading hypothetical, described health states [7].

Direct TTO and the EQ-5D descriptive system thus enjoy different reference populations, as respondents of direct TTO act as their own reference and express their own preferences. Hence, when both methods are applied to a set of actual patients, the instruments are conveying different views or sets of preferences [6,15]. Direct TTO may be seen as reflecting the individual’s preferences regarding his or her own health state, whilst the EQ-5D descriptive system instead reflects the preferences of healthy individuals evaluating hypothetical, described health states. Previous research has pointed out potential discrepancies between the outcomes of the two methods [3,36].

Differing outcomes from the two methods could be of large importance since it may give different notions of health impairment due to choice of method. Allocation of resources could thereby seem improper depending on assessment method applied. At worst, it could get to the point that treatments would be subsidized if using one method for assessment of QoL but not when using another. Thus, it is motivated to investigate such a possible difference in outcomes. If outcomes differ, a discussion regarding which method to apply could also be motivated.
2 Study Objectives

The objectives of this study were to:

I. Compare how patients with an osteoporosis-related fracture estimate their quality of life with the EQ-5D descriptive system and directly elicited Time-Trade-Off grading respectively, assessing if these instruments provide a difference in outcomes.

II. Investigate what the implications may be using each method respectively, in terms of a cost-utility analysis.
3 Background

3.1 Medical Background

Osteoporosis is a medical condition characterized by low bone mass as well as impaired micro architecture of the bone tissue. This leads to a more fragile bone structure in afflicted individuals, enhancing the occurrence of fractures – even from low-energy trauma. Diagnosing patients with osteoporosis is therefore needed to determine their risk of receiving a fracture and to enable appropriate prevention and treatment. The diagnosis is based upon the patients’ bone mineral density (BMD) in relation to that of young, healthy individuals. WHO proposes the definition of osteoporosis to refer to individuals with a BMD of \( \geq 2.5 \) SD below the mean BMD in young, healthy individuals of the same population [25].

3.2 Economic Evaluation

During the last two decades, governmental agencies with aims of determining allocation policies have been established in a number of countries, e.g. National Institute of Clinical Excellence in England and Wales (NICE, 1999), at the Commonwealth Department of Human Services and Health in Australia (1992) and at the Ministry of Health Ontario in Canada (1994) [18,32]. These agencies have specific requirements for submission of economic evaluations regarding medical products, and as the trend goes towards assessing how to allocate resources efficiently, cost-effectiveness analyses are becoming a common way of comparing therapies and drugs [18].

CEAs highlight how big the health benefits are with an intervention compared to its costs. There is a distinction made between CEAs comparing intervention-specific measures and the type of CEA performed with QoL as factor for comparison. The latter is also called cost-utility analysis (CUA) since it only comprises changes in health-related utility in the denominator (see below). The denominator hence consists of accumulated changes in QoL over time, usually expressed in terms of Quality-Adjusted Life-Years [18].
3.2.1 Quality-Adjusted Life-Years

The Quality-Adjusted Life-Year (QALY) is a measure to determine cumulative QoL over time. Assuming no discounting, it can be computed as follows [8]:

\[
\text{QALY} = H \times Y
\]

where \( H \) represents the quality weight of the health state, i.e. the health-related utility, and \( Y \) represents the time in years spent in that health state [8].

The QALY hence reflects quantity of life as well as perceived quality of life [11,12,15]. The computation can also be visualized by the area under the curve in Figure 1. As the QALY is a common unit of measure irrespective of medical condition, it enables comparison of costs and QoL implications associated with therapies concerning different diseases [30].

There are the axioms of expected utility theory to consider in order for quality weights and QALYs to be seen as utilities. The QALY should not be seen as a utility measure by conventional means. However, as Drummond argues [11], the QALY may still be a useful measure in practice since it approximates a measure for what health care aims at maximizing [11]. The von Neumann’s and Morgenstern’s axioms of expected utility theory are elaborately described in literature on the subject[1,11] and will not be further investigated here.

3.2.2 The Incremental Cost-Effectiveness Ratio

The comparison of e.g. two drugs A and B through CUA is possible to perform through employing QALYs to compute an incremental cost-effectiveness ratio (ICER), as follows [18]:

\[
\text{ICER} = \frac{\text{Cost of drug A} - \text{cost of drug B}}{\text{QALY gain with drug A} - \text{QALY gain with drug B}}
\]
As seen above, the ICER is always a relative measure, comparing two alternatives of drugs or therapies. When determining whether or not to subsidize a drug, authorities generally apply a cost-effectiveness threshold concerning how much they are actually willing to pay per QALY gained due to usage of the drug [29]. Even though objections towards cost-effectiveness thresholds have been put forward in the literature [11], national authorities, e.g. NICE in the UK, practice such a threshold which hence embodies society’s willingness to pay (WTP) for an intervention [22]. The Dental and Pharmaceutical Benefits Agency in Sweden (TLV) estimates society’s WTP to approximately SEK 600,000 per QALY gained, reflecting which drugs to be included in the pharmaceutical benefits in Sweden. This is not always the case in practice, but a goal for aiming at in theory. Specifically, exceptions from the regular cost-effectiveness threshold may be carried out for extraordinarily ill patients [29].

To investigate the importance of potential differences in instrument outcomes, it is of interest to have a look at the practical implications. This can be done through examining how an ICER appears based on outcomes from the two different methods for assessing QoL respectively, i.e. outcomes from direct TTO versus outcomes from the EQ-5D.
3.3 Assessing and estimating Quality of Life

3.3.1 Direct Time-Trade-Off grading

Direct TTO is based on an individual’s choice of staying in either a state of impaired health for a certain amount of time \( t \) or a state of full health for a shorter time period \( x \) of time \( t \), both alternatives followed by immediate death (Figure 1). The state of impaired health is given a utility index with a value depending on how large amount of time in full health is needed for the individual to be indifferent between the two alternatives. The health utility index value for the impaired health state amounts to \( x/t \), as the individual is indifferent between \( x \) years in full health and \( t \) years in a chosen state of impaired health [15]. The time horizon \( t \) for direct TTO grading in the study performed to generate the EQ-5D descriptive system was set to 10 years [7]. The method of direct TTO discussed in this paper considers chronic health states, as given by the statement of immediate death after the time horizon \( x \) or \( t \) [11].

![Figure 1. Overview of the preference-based technique used for Time-Trade-Off grading of chronic health states.](image)

In a pilot study, Torrance[31] determined the method of direct TTO to be the most easily administered method for the purpose of evaluating a state of health [31].
3.3.2 EQ-5D descriptive system

Direct TTO has, after its adaption into practical use, become the basis of a generic instrument for assessing QoL in individuals; the EQ-5D descriptive system (see appendix A) [7]. In the EQ-5D, daily life is subdivided into five dimensions (5D); 1) mobility, 2) self-care, 3) usual activities, 4) pain/discomfort and 5) anxiety/depression [7]. The respondent grades the five dimensions on a scale of three levels – whether the respondent has got no problems, moderate problems or severe problems with each dimension respectively. These levels of severity are subjectively chosen by the patient for each of the five dimensions. Since the three levels are not measured using a numerical scale, this approach implies that the equivalent numerals 1-3 of the three levels are only of ordinal nature and may not be used as cardinal scores [27]. Thus, each of the three levels has a clear ranking compared to the others whilst one level by itself is not alone associated with a utility value. A cardinal score or utility on the other hand implies a certain value by itself; a number is attached, translating its value or utility [11]. The three EQ-5D levels are therefore only to be used for internal ranking compared to one another amongst them and may not be assigned any utility values.

Filling out the EQ-5D descriptive system implies the respondent’s score to be one out of 243 possible health states, since there are $3^5 = 243$ possible outcomes (states of unconsciousness and death not included) [7,28]. To then evaluate patients’ different health states and their impairment on life-quality, it is necessary to assign health utility index values – quality weights – to each health state [7,27]. The quality weights assigned then enable the computation of QALYs [15].

In 1997, Dolan [8] performed a study to designate a utility value to each possible health state, making it possible to assign each of the 243 health states an index value of maximally 1 and with a value of 0 equivalent to being dead [7]. Utility values were assigned through direct TTO grading of all health states evaluated. Direct valuations were not generated for each and every health state, albeit for 42 of them. Values for the remaining health states were then interpolated, since direct valuations of all 243 states would require either a major cohort of respondents or a significant amount of time engaged for each respondent. Interpolation was possible since the respondents also did arrange the medical conditions in order according to preferences of which health state was perceived more favorable than the other [7]. Since
patients were asked to define the proposed hypothetical health states as better or worse than being dead in the study by Dolan [7], some health states were through a special version of TTO-grading\(^1\) assigned a negative QoL value [7,9]. Health states evaluated through EQ-5D may also be described in terms of a five-digit number corresponding to grading of the five dimensions, e.g. 11312 corresponding to no problems with either mobility, self-care or pain/discomfort, although moderate problems with anxiety/depression and extreme problems with usual activities [27].

The matter of particular interest when Dolan assigned health utility index values for EQ-5D health states is the fact that the people evaluating the different health states were not in fact afflicted by impaired health – the respondents were only evaluating hypothetical, described health states as the health states were not a reality for the respondents themselves. Since evaluated health states were only hypothetical, the respondents assigned utilities based on beliefs of how they would perceive the health states in question. The EQ-5D descriptive system would be a method motivated to use if its translated outcomes are equivalent to the outcomes of direct TTO [36].

3.4 Previous research

There are studies on potential empirical differences between the outcomes of the EQ-5D descriptive system and direct TTO, such as Burström et al [3] from 2006 and Johannesson & Zethraeus [36] from 1999, although neither with a multinational study sample as is the case in this particular study. This study is also partially new in its type since QoL estimation of health states was performed during interviews instead of letting patients grade their perceived QoL on their own (as in Burström et al [3]). Burström et al [3] and Johannesson & Zethraeus [36] both suggest that instrument outcomes differ, and that the difference may be even larger for conditions of more severe health impairment. These two studies were based on samples of individuals grading their own health states; a general population survey (n=2,549) [3] and a sample of menopausal women under hormone-replacement therapy (n=104) [36] respectively.

\(^1\) Health states earlier defined as worse than death received an index value of \((-x/(10-x))\) [7].
Nation-specific validation studies of the relation between EQ-5D values and values from direct TTO have been performed – all considering hypothetical, described health states graded by population samples of healthy individuals. These validation studies have all had the purpose of creating a local, nation-specific, EQ-5D algorithm with respect to the locally prevailing preferences [13,23,28]. The national value sets available have earlier been compared in descriptive analyses performed by the EuroQol Group and others [17,28].

In previous research, a distinction is made regarding the two ways of evaluating health states – the direct time-trade-off grading vs. the indirect time-trade-off grading made by filling out the EQ-5D descriptive system. The EQ-5D method has been referred to as being prevailed by social preferences, as the technique for evaluating health states that way is based upon individuals from the general public in society and who are not afflicted by the medical condition in question [36], even though these preferences are not related to altruism in any sense. Directly elicited time-trade-off grading, on the other hand, is by definition based on the individual patient's preferences regarding his or her health state, in line with the individualistic base of welfare economics [15]. The welfare economists assume that individuals maximize a well-defined utility function [12,15], i.e. that no one may better judge an individual's welfare than the individual herself [2]. Arguments have however been put forward in favor of both approaches mentioned above [15].
4 Data

To analyze outcomes from the two instruments, a prospectively collected cohort of osteoporosis patients suffering from an osteoporosis-related fracture was used. Patient-level data were accessed by means of The International Costs and Utilities Related to Osteoporotic fractures Study (ICUROS), which’s study design and procedure were approved by ethical committees in the countries concerned.

4.1 ICUROS study design

The material for disposal has been continuously collected since 2007, at 42 clinical centers in eleven different countries. Through interviews performed by nurses, patient information was collected either face-to-face (for inpatient-treated during first contact) or via telephone (outpatient-treated during first contact and for all patients during subsequent contacts). The primary material consisted of patient-level data from 4320 patients regarding baseline characteristics, costs related to the fracture and estimated quality of life at different points in time after the fracture event. At first contact patients estimated their QoL regarding time just before and just after fracture (two estimates), implying that the utility estimates assigned for the time before fracture were retrospectively collected. Then, interviews were performed at 4 months, 12 months and 18 months after fracture, utility estimates being collected prospectively.

Patients were enrolled in the study as they sought medical care at one of the clinical centers included in the study, being diagnosed with either a hip fracture, wrist fracture or a vertebral, humeral or ankle fracture. To create a multinational sample of patients, study centers were chosen in different countries [14]. The health status of enrolled patients in terms of utility estimates was assessed using the EQ-5D descriptive system and a time-trade-off question (see appendices). These health status assessments through formularies were performed during interviews with health personnel.

Data relevant to this specific study were provided by the sole patient, hence not relying on patient details from registries provided by authorities or other institutions. Health personnel
at the local study centre were then responsible for import of patient information into the web-based registry.

### 4.2 Patient selection

Data extractions were performed from the web-based registry of ICUROS in January 2011. Specific inclusion and exclusion criteria for the ICUROS study can be found in appendix C.

![Timeline for data collection in the ICUROS study.](image)

#### 4.2.1 Statistical comparison of mean outcomes

For performance of statistical comparisons of mean outcomes, it was desirable to use the same subsample over the different time points for evaluation. Lower limit for inclusion in these analyses was set to 12 months’ follow-up, enhancing comparisons of QoL estimates over three different time points for estimation. Due to the relatively small amount of patients having yet finished the whole ICUROS study process of 18 months, this last time point for QoL estimation was not taken into account in these analyses (Figure 2).

#### 4.2.2 Estimating an EQ-5D value set

On the contrary, no upper or lower limit was set regarding time length of patients’ participation in the study when estimating an EQ-5D value set to be based on direct TTO grading from actual patients. Consequently, all patients in the study sample were included for
the value set estimation. This was desirable in order to make the patient material as large as possible and hence the estimated value set as trustworthy as possible. Furthermore, including all observations was motivated since the aim with estimating a value set was to assess the link between TTO values and EQ-5D scores; it was hence assumed that the individual performs her direct TTO in relation to her EQ-5D scores no matter which point in time after having received a fracture it may be. The TTO value should reflect the individual preferences irrespective of where in the course of a disease an individual lies.

4.2.3 Practical implications in Cost-Utility Analysis

Patients having received either a humeral or an ankle fracture were excluded when simulating performance of a cost-utility analysis, as these fracture types were not included in the cost-utility model used. Patient selection for the simulated CUA was based on the subset of patients having participated in the ICUROS study for 12 months’ time or more (Figure 2), since one year’s QoL estimates were needed in order to compute utility multipliers required (see Methods II below).
5 Methods

Statistical software used was MS Access 2003 and MS Excel 2003 (Microsoft Corporation) for data management as well as STATA 10 (StataCorp LP, College Station, TX) for further data management and execution of statistical analyses.

When analyzing QoL data, negative values of QoL were not rounded off to zero, thus kept as negative values. Since the design of the EQ-5D value sets enable negative values, it is adequate to keep this spectrum; the sample would be arbitrarily skewed if values were allowed to be rounded off. The original EQ-5D algorithm from Dolan [7] was used to calculate utility values from the EQ-5D scores, enabling comparison with the values from direct TTO. Direct TTO values (see appendix B) were divided by ten to obtain values ranging from 0 to 1, in order to enhance comparisons.

QoL values are time-specific in the sense that they convey the patient’s perceived life-quality at a specific point in time (e.g. 4 months or 12 months after fracture). An exception in this case is the QoL baseline estimates – regarding time before fracture – which were recollected during the first interview. The analyzed scenario comprises the assumption that the QoL level assigned at four months after fracture was also attained at four months after fracture, linearly improved from the time of fracture.
5.1 Methods I: Comparison of instrument outcomes

In line with the first study objective, outcomes from the two instruments were compared. Comparisons were performed in two ways with different approaches; through statistical comparisons of mean instrument outcomes, and through estimation of an EQ-5D value set based on data from the ICUROS study regarding direct TTO and scores from the EQ-5D descriptive system. The estimated value set was aimed at being compared to the original value set.

5.1.1 Statistical comparison of mean outcomes

Paired sample t-tests of the two instruments’ mean outcomes were performed to assess whether they provided a statistically significant difference, sorted by fracture type and time point for QoL estimation respectively (at time of fracture, after 4 months and after 12 months, \( p<0.05 \)). Also, computations of confidence intervals (95%) for the mean outcomes were performed, sorted by fracture type as well since the mean outcomes empirically showed to differ depending on which fracture type was subject to evaluation [21]. The sample of differences between instrument outcomes was assumed to be normally distributed with regard to the Central Limit Theorem\(^2\), enabling performance of confidence intervals according to the student's t distribution [21]. Since the variance of the population means (of direct TTO and EQ-5D outcomes respectively) are not known, the t distribution was used to take into account the possible large spread of population means. However, a non-parametric test was also executed to assess the differences in outcomes, for matched pairs since the samples consisted of the same patients performing evaluations using two different methods. Mean values of instrument outcomes were compared through the Wilcoxon signed rank test. The sample was approximated to be normally distributed under the null hypothesis regarding the Wilcoxon test since the sample size exceeded 20 observations [21]. Null hypothesis stated that the differences in outcomes were centered on zero, i.e. that there was no differences in mean instrument outcomes. Alternative hypothesis stated the opposite, i.e. \( \text{HA: difference in mean outcomes } \neq 0, \ p<0.05 \).

\(^2\) The Central Limit Theorem suggests that the distribution of sample means from uniform or symmetric distributions may be closely approximated by the normal distribution – generally for samples exceeding 25 observations [21].
5.1.2 Estimating an EQ-5D value set

Estimating a new value set was appropriate since it enabled a systematic comparison with the original value set, hence enabling a comparison of how health state evaluation depends on whether evaluated health states are actually experienced by the evaluator or only hypothetical and solely described to the evaluator. Thus, a difference in perceptions between the individual patients and the healthy population would be further illuminated.

Regression analyses were performed with values from direct TTO as predicted variable, and using dummy variables for levels 2 and 3 graded on the five dimensions of EQ-5D as explanatory variables. An N3 term, used in the original value set by Dolan [7], was also included to enhance comparison. The N3 term is a dummy variable taking on a value of 1 if the patient grades the lowest out of the three EQ-5D levels in any one dimension. Since the three levels of each dimension are not to be regarded as cardinal scores, they were assigned dummy variables when performing regression analyses to enable correct interpretation of the coefficient estimates. These coefficient estimates – set of values – were then supposed to correspond to the health-related utilities associated with the different levels graded within each dimension. Direct TTO values were thus considered a surrogate variable for perceived QoL.

Values from direct TTO were divided by ten in order to obtain a value in the range from 0 to 1, as equivalent to the levels of the EQ-5D values. Thereby, it was possible to receive coefficient estimates on equivalent levels as in previous research when performing regression analyses. This is also the reason to the name of the predicted variable (tto_dec) in below presentation of regression analyses. The dataset consisting of the ICUROS patient-level data used was declared to contain panel data, longitudinal data, since having both a time series and a cross-sectional dimension [35].

Estimations performed by patients regarding their perceived QoL before actually having the fracture were excluded due to the risk of recollection bias entailed. Recollection bias is here
referring to the act of not remembering the previous health state in objective terms in relation to the post-fracture condition just entered.

When performing regression analyses, a generalized least-squares technique managing the error terms with the random-effects method was used as well as managing the error terms with the fixed-effects method [35]. A Hausman test was also performed to assess whether the two methods of managing the error terms differed systematically or not, i.e. whether the question of which method to use was even relevant at all. However, both the fixed-effects and the random-effects methods were applied to enable comparison.

Regression analyses performed were configured as shown below:

\[ tto_{\text{dec}} = \beta_0 + \beta_{\text{mob2}} \cdot \text{mob2} + \beta_{\text{mob3}} \cdot \text{mob3} + \beta_{\text{infl2}} \cdot \text{selfcare2} + \beta_{\text{infl3}} \cdot \text{selfcare3} + \beta_{\text{usualact2}} \cdot \text{usualact2} + \beta_{\text{usualact3}} \cdot \text{usualact3} + \beta_{\text{pain2}} \cdot \text{paindis2} + \beta_{\text{pain3}} \cdot \text{paindis3} + \beta_{\text{anxdep2}} \cdot \text{anxdep2} + \beta_{\text{anxdep3}} \cdot \text{anxdep3} + \beta_N \cdot N \]

Since dummy variables were given for levels 2 and 3 of each dimension, the benchmark health state referred to “no problems” (level 1) regarding all five dimensions. Consequently, the dependent variable, i.e. health-related utility, was defined as \( 1 - X \) where \( X \) represents the value assigned to a particular health state.

To further enhance practical comparison of the estimated value sets to the original value set, three examples of EQ-5D health states were assigned utility values according to the estimated value sets and the original value set respectively.
5.2 Methods II: Practical implications in Cost-Utility Analysis

For the purpose of simulating a cost-utility analysis, a model that has been published in previous health-economical research [16] was used, designed as shown in Figure 3 below.

![Markov model with possible health states and transitions visualized](image)

Figure 3. A Markov model with possible health states and transitions visualized. After having received either a hip or a vertebral fracture, the individual goes into a post-fracture state. Death is always a possible health state. Model structure published with consent from authors of original study [16].

The model is a Markov model which refers to that spending the time of one cycle in one of the given health states is associated with a defined utility and a defined cost entailed [18]. This model was initially constructed for modeling different treatments for osteoporosis regarding their costs and improvement in quality of life, previously published by Jönsson et al [16]. Probabilities of health state transitions (incidences) and other figures such as mortality rates from Jönsson et al [16] were used, obtained from registries declared in the previously published study [16]. The above model was designed with a cycle length of 6 months, although fracture states were maintained for 12 months. The model had a time horizon of 5 years, meaning that individuals of the cohort are followed for 5 years or to the point of dying or becoming 100 years old.

The cohort analyzed in the model was assumed to have a T-score of ≥ 2.5 standard deviations below the score of young, healthy women, as WHO’s definition of osteoporosis [16,33]. The individuals of the cohort were further assumed to have a mean age of 71 years.
when starting their osteoporosis treatment since this is the average age for Swedish postmenopausal women at the time for starting their osteoporosis treatment [20]. Adherence to treatment in the model was based on an earlier performed Swedish study of persistence to osteoporosis treatment [26]. The two treatment options compared in this specific case of running the model (with specific probabilities for transitions between health states depending on medication) were a hypothetic drug versus no treatment. Since the aim of performing a cost-utility analysis was to investigate implications of using different value bases for health-related utilities, comparing a hypothetic drug to no treatment was considered sufficient. The hypothetical drug evaluated through this model was assumed to be superior to no treatment by a relative risk reduction regarding future fractures, affecting quality-of-life aspects and health-care costs avoided as well as mortality rates due to fracture avoidance. Further description and details of this specific model can be obtained from the article on the subject by Jónsson et al [16].

Since the health states must be assigned utilities entailed to enable cost-utility analysis, a utility multiplier for each health state must be computed. The utility multiplier is state-specific and determines the utility associated with being in a particular health state, regardless of at what point in time a transition to the health state takes place. Utility multipliers were computed as follows:

\[
\text{Utility multiplier} = \frac{\text{QoL if having a fracture}}{\text{QoL if not having a fracture}} = 1 - \frac{\text{QoL loss in impaired health state}}{\text{QoL before fracture}}
\]

The utility multiplier is supposed to show how large is the perceived utility of a health state over time, thus based on the cumulative quality-of-life loss during a certain amount of time, e.g. a year. When computing utility multipliers, the estimated values for health utility before fracture were assumed to be maintained during the following year if the fracture had not occurred, enabling computation of mean QoL loss due to fracture. Patient characteristics
from the ICUROS study data were not applied in the CUA, only were the utility values observed for predicting the utility multipliers used in the model regarding the first year after fracture. Model settings as in the study by Jönsson et al [16] were used, although using costs in Swedish kronor (SEK) in this study. Utility multiplier for the state of “other osteoporotic fracture” was imported from Jönsson et al [16], since it was based on fractures of the pelvis, rib, humerus, clavicle, scapula, sternum etc. [16]. Hence, an approximation of the health-related utility for such a composite health state in the model used was not possible to compute through the ICUROS data set.

In this study, the model was used only for computing ICERs and QALY gains associated with a hypothetical drug compared to no treatment – therefore not proposed for further interpretations. The model was intended to demonstrate what implications the different methods for estimating QoL may have on a typical CUA and thereby on ICERs attained with a model of this kind. Computation of ICERs depending on method for QoL estimation was obtained through using the two different sets of utility multipliers in the model, one at a time. Thus, the only thing altered between computing the two ICERs was the utility multiplier set regarding hip, vertebral and wrist fractures concerning the first year after a fracture event. The utility multipliers regarding EQ-5D scores were computed based on the original EQ-5D value set from Dolan [7].
6 Results

6.1 Results I: Comparison of instrument outcomes

Below are the results of analyses regarding instrument outcomes, i.e. QoL estimates for patients from the ICUROS study.

6.1.1 Statistical comparison of mean outcomes

Table 1 shows baseline characteristics for the subset of patients used in comparison of mean outcomes from the two instruments.

Table 1. Baseline characteristics for subset of patients used to perform statistical comparisons of mean instrument outcomes.

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean age</th>
<th>Males</th>
<th>Females</th>
<th>Hip</th>
<th>Wrist</th>
<th>Vertebral</th>
<th>Humeral</th>
<th>Ankle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>72</td>
<td>25%</td>
<td>75%</td>
<td>64</td>
<td>39</td>
<td>28</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Spain</td>
<td>79</td>
<td>22%</td>
<td>78%</td>
<td>39</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>69</td>
<td>15%</td>
<td>85%</td>
<td>51</td>
<td>47</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Italy</td>
<td>76</td>
<td>17%</td>
<td>83%</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lithuania</td>
<td>74</td>
<td>33%</td>
<td>67%</td>
<td>21</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mexico</td>
<td>92</td>
<td>0%</td>
<td>100%</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Russia</td>
<td>65</td>
<td>16%</td>
<td>84%</td>
<td>119</td>
<td>200</td>
<td>170</td>
<td>105</td>
<td>161</td>
</tr>
<tr>
<td>ALL</td>
<td>67</td>
<td>18%</td>
<td>82%</td>
<td>297</td>
<td>297</td>
<td>225</td>
<td>130</td>
<td>162</td>
</tr>
</tbody>
</table>
Table 2. Overview of mean estimated Quality of Life with EQ-5D and directly elicited TTO respectively, sorted by fracture type and time after fracture event.

<table>
<thead>
<tr>
<th></th>
<th>Time point</th>
<th>TTO decimal (95% CI)</th>
<th>EQ-5D (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hip fracture</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before fracture</td>
<td>0.85</td>
<td>(0.83-0.88)</td>
<td>0.78</td>
</tr>
<tr>
<td>0 months</td>
<td>0.63</td>
<td>(0.60-0.66)</td>
<td>-0.07</td>
</tr>
<tr>
<td>4 months</td>
<td>0.77</td>
<td>(0.74-0.80)</td>
<td>0.50</td>
</tr>
<tr>
<td>12 months</td>
<td>0.74</td>
<td>(0.71-0.77)</td>
<td>0.60</td>
</tr>
<tr>
<td><strong>Wrist fracture</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before fracture</td>
<td>0.95</td>
<td>(0.93-0.97)</td>
<td>0.87</td>
</tr>
<tr>
<td>0 months</td>
<td>0.75</td>
<td>(0.72-0.78)</td>
<td>0.44</td>
</tr>
<tr>
<td>4 months</td>
<td>0.93</td>
<td>(0.91-0.95)</td>
<td>0.79</td>
</tr>
<tr>
<td>12 months</td>
<td>0.93</td>
<td>(0.91-0.95)</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Vertebral fracture</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before fracture</td>
<td>0.91</td>
<td>(0.89-0.93)</td>
<td>0.81</td>
</tr>
<tr>
<td>0 months</td>
<td>0.44</td>
<td>(0.39-0.48)</td>
<td>0.23</td>
</tr>
<tr>
<td>4 months</td>
<td>0.88</td>
<td>(0.86-0.90)</td>
<td>0.69</td>
</tr>
<tr>
<td>12 months</td>
<td>0.89</td>
<td>(0.86-0.92)</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Humeral fracture</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before fracture</td>
<td>0.90</td>
<td>(0.87-0.92)</td>
<td>0.84</td>
</tr>
<tr>
<td>0 months</td>
<td>0.55</td>
<td>(0.51-0.60)</td>
<td>0.37</td>
</tr>
<tr>
<td>4 months</td>
<td>0.88</td>
<td>(0.85-0.92)</td>
<td>0.77</td>
</tr>
<tr>
<td>12 months</td>
<td>0.90</td>
<td>(0.86-0.93)</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Ankle fracture</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before fracture</td>
<td>0.95</td>
<td>(0.93-0.97)</td>
<td>0.86</td>
</tr>
<tr>
<td>0 months</td>
<td>0.54</td>
<td>(0.49-0.58)</td>
<td>0.20</td>
</tr>
<tr>
<td>4 months</td>
<td>0.90</td>
<td>(0.87-0.94)</td>
<td>0.77</td>
</tr>
<tr>
<td>12 months</td>
<td>0.90</td>
<td>(0.86-0.94)</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 2 offers an overview of mean outcomes from the two instruments as well as confidence intervals computed. 0 months refers to time of the fracture event.

Paired sample t-tests of the sample means showed statistically significant differences between the instruments’ mean outcomes. Comparing the different outcomes by fracture type, there were statistically significant differences ($p<0.05$) between the EQ-5D mean...
estimate and the TTO mean estimate at all time points. The EQ-5D values were consistently lower than the values from directly elicited time-trade-off grading, regardless of fracture type.

The Wilcoxon signed rank test for paired samples was performed, null hypothesis stating that the differences between instrument outcomes were centered on zero. The null hypothesis was rejected, hence showing that the outcomes of the two instruments were of statistically significant difference, for all five fracture types and all four time points respectively.

6.1.2 Estimating an EQ-5D value set

In Table 3 the estimated EQ-5D value set is depicted. Statistical significance for coefficients was received to a larger extent when using the random-effects method than when using the fixed-effects method. Performance of Hausman’s test showed that the coefficient estimates systematically differed between these two methods of managing the error terms ($p<0.05$), i.e. the null hypothesis that there were no systematic differences ($H_0: \text{Cov(coef with FE, coef with RE)} \neq 0$) was rejected. Results from both methods are depicted below.

Below is an overview of coefficient estimates from the different regression analyses performed as well as from the original algorithm to enable comparison. Table 4 depicts three different EQ-5D health states and their assigned quality weights depending on value set used.
Table 3. Overview of EQ-5D value sets estimated in this study together with the original value set.

<table>
<thead>
<tr>
<th>Mobility</th>
<th>Fixed effects</th>
<th>Random effects</th>
<th>Original algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.934*</td>
<td>0.951*</td>
<td>0.919*</td>
</tr>
<tr>
<td>Mobility level 2</td>
<td>0.025*</td>
<td>-0.039*</td>
<td>-0.069*</td>
</tr>
<tr>
<td>Mobility level 3</td>
<td>-0.002</td>
<td>-0.117*</td>
<td>-0.314*</td>
</tr>
<tr>
<td>Self-care level 2</td>
<td>-0.165*</td>
<td>-0.164*</td>
<td>-0.104*</td>
</tr>
<tr>
<td>Self-care level 3</td>
<td>-0.142*</td>
<td>-0.166*</td>
<td>-0.214</td>
</tr>
<tr>
<td>Usual activity level 2</td>
<td>-0.020</td>
<td>-0.011</td>
<td>-0.036*</td>
</tr>
<tr>
<td>Usual activity level 3</td>
<td>-0.034</td>
<td>-0.036*</td>
<td>-0.094*</td>
</tr>
<tr>
<td>Pain/discomfort level 2</td>
<td>-0.059*</td>
<td>-0.029*</td>
<td>-0.123*</td>
</tr>
<tr>
<td>Pain/discomfort level 3</td>
<td>-0.210*</td>
<td>-0.131*</td>
<td>-0.386*</td>
</tr>
<tr>
<td>Anxiety/depr level 2</td>
<td>-0.079*</td>
<td>-0.076*</td>
<td>-0.071*</td>
</tr>
<tr>
<td>Anxiety/depr level 3</td>
<td>-0.062*</td>
<td>-0.151*</td>
<td>-0.236*</td>
</tr>
<tr>
<td>N3</td>
<td>0.019</td>
<td>0.059*</td>
<td>-0.269*</td>
</tr>
<tr>
<td>Number of observations</td>
<td>7,447</td>
<td>7,447</td>
<td>35,964</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.25</td>
<td>0.29</td>
<td>0.46</td>
</tr>
</tbody>
</table>

* = p<0.05

Table 4. Examples of EQ-5D health states and their quality weights depending on method used for QoL estimation; value sets estimated within this study compared to the original algorithm.

<table>
<thead>
<tr>
<th>Health State</th>
<th>Fixed effects</th>
<th>Random effects</th>
<th>Original algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>12111</td>
<td>0.769</td>
<td>0.787</td>
<td>0.815</td>
</tr>
<tr>
<td>22123</td>
<td>0.692</td>
<td>0.627</td>
<td>0.118</td>
</tr>
<tr>
<td>23233</td>
<td>0.500</td>
<td>0.509</td>
<td>-0.371</td>
</tr>
</tbody>
</table>
6.2 Results II: Practical implications in Cost-Utility Analysis

*Figure 4* depicts the exclusion of patients performed from the initial patient cohort. In total, 819 individual patients’ data were used for simulating a cost-utility analysis.

![Flow chart showing exclusion of patients from collected sample before computation of utility multipliers and simulation of cost-utility analysis were performed.](image)

In *Table 5* below, figures concerning patients that were included in the simulated cost-utility analysis are shown.
Table 5. Baseline characteristics for subset of patients whose utility values were used in computation of utility multipliers and simulation of cost-utility analysis.

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean age</th>
<th>Males</th>
<th>Females</th>
<th>Hip</th>
<th>Wrist</th>
<th>Vertebro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>72</td>
<td>25%</td>
<td>75%</td>
<td>64</td>
<td>39</td>
<td>28</td>
</tr>
<tr>
<td>Spain</td>
<td>79</td>
<td>22%</td>
<td>78%</td>
<td>39</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>69</td>
<td>15%</td>
<td>85%</td>
<td>51</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>Italy</td>
<td>76</td>
<td>17%</td>
<td>83%</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Lithuania</td>
<td>74</td>
<td>33%</td>
<td>67%</td>
<td>21</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Mexico</td>
<td>92</td>
<td>0%</td>
<td>100%</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Russia</td>
<td>65</td>
<td>16%</td>
<td>84%</td>
<td>119</td>
<td>200</td>
<td>170</td>
</tr>
<tr>
<td>All</td>
<td>68</td>
<td>18%</td>
<td>82%</td>
<td>297</td>
<td>297</td>
<td>225</td>
</tr>
</tbody>
</table>

Utility multipliers computed from the selected subsample are shown in Table 6 below.

Table 6. Utility multipliers for the Markov states in the model

<table>
<thead>
<tr>
<th></th>
<th>Hip fracture</th>
<th>Wrist fracture</th>
<th>Vertebro fracture</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTO</td>
<td>0.86</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.56</td>
<td>0.86</td>
<td>0.77</td>
</tr>
</tbody>
</table>

The utility multipliers point out a higher perceived QoL having a wrist fracture compared to having either a hip fracture or a vertebral fracture. Patients having received a hip fracture perceived a greater health-related utility impairment compared to patients with other fracture types.

Table 7. Incremental QALY gain with hypothetic drug compared to no treatment as well as ICERs for hypothetic drug versus no treatment, computed on outcomes from both QoL estimation instruments, model time horizon set to five years.

<table>
<thead>
<tr>
<th></th>
<th>Incremental QALY gain</th>
<th>SEK per incremental QALY gained</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTO</td>
<td>0.066</td>
<td>205 242</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.074</td>
<td>183 479</td>
</tr>
</tbody>
</table>

In Table 7, ICERs and incremental QALY gains are given for a hypothetic drug compared to no treatment of osteoporosis patients, based on outcomes from the two different instruments. Hence, the only factor differing between the two ICERs above is the way of assessing patients’ perceived QoL. In this specific case, the differences amounted to an approximately 12 percent higher cost per QALY with TTO based QoL values compared to when using EQ-5D values. The QALY gains with the hypothetic drug compared to no treatment amounted to a 12 percent higher value when based on EQ-5D values compared to when using values from direct TTO. These figures concern a model time horizon of five years.
7 Discussion

Since the EQ-5D method for assessing Quality of Life was initially based upon the concept of Time-Trade-Off grading, a comparison of the two instruments outcomes has been motivated. In line with previous research, this study suggests that the instrument outcomes from the EQ-5D descriptive system and directly elicited time-trade-off grading consistently differ, when studying a cohort of patients who grade their own health statuses. QoL estimates acquired with the EQ-5D descriptive system’s original algorithm are underestimating the level of perceived quality of life compared to the estimates received from directly elicited TTO grading. In line with previous research, outcome differences between the two instruments grow larger when more severe health states are subject to evaluation.

Values of the EQ-5D descriptive system are given by healthy individuals’ perception of how the experience of a medical condition would appear, whilst direct TTO grading in the ICUROS study regarded evaluation of health states that are actually experienced by the respondents. Disparities in instrument outcomes from the two methods evaluated could be derived from the different views of a health state depending on being in it or merely hypothesizing upon it. Apparently, evaluation of a described health state implies a lower estimated QoL than evaluation of an actually experienced health state does. Having entered an impaired state of health and coping with it may be major explanations to the consistently differing outcomes from healthy people’s QoL grading compared to actual patients’ QoL grading of equivalent health states. In other words, having de facto become ill could imply an acceptance of the condition and an overall appreciation of e.g. being alive.

The extent to which outcomes differ depending on method seems to rely on severity of health state evaluated. Thus, certain levels of some dimensions imply larger differences between the view of the actual patient and the view of the speculating, healthy individual. This can be further illustrated by the hip fracture state, which gives the largest relative difference in mean utility estimates depending on method out of the five fracture types evaluated in this thesis. As differences between the original algorithm and the algorithm estimated based on the ICUROS data appear, larger relative discrepancies for the hip fracture state compared to other fracture states may be a result of that a hip fracture
comprises restraints for several of the dimensions in which the coefficient estimates distinctively differ between value sets compared, e.g. regarding pain/discomfort and mobility. As EQ-5D’s original algorithm offers a broader spectrum of utility estimates (from -1 to 1) than direct TTO does (0 to 1), the accentuated difference in outcomes with increased severity is an inevitable consequence.

In cost-utility evaluations of medical interventions the QALY gains are supposed to be comprehensible, computed through a generic measure. However, as this thesis has tried to illuminate, the quality weights assigned may differ significantly depending on method used for QoL estimation. Such differences may affect the decisions regarding resource allocation, and is therefore a matter of importance to be aware of. As seen when simulating a cost-utility analysis using utilities from either EQ-5D or direct TTO grading, it is clear that the practical implications of method for QoL estimation are notable. Since a difference in QALY gains is received depending on method used for QoL estimation, the incremental cost-effectiveness ratios also differ. Even though this is a highly hypothetical cost-utility analysis being performed, the different ICERs illustrate an important implication of methods used for QoL estimation. In the extreme case, the subsidization of a drug could rely on the choice of method used for QoL estimation when the drug is evaluated.

Estimating a value set with the random-effects method implies an error term with two components – one that is unique to each observation and one that represents the deviation due to the individual respondent. Consequently, the random-effects (RE) method allows for differences between individuals as well as differences within individuals between time points. The RE method could hence be considered the most appropriate. The N3 variable was positive, and statistically significant, using the RE method, which may seem contradictory. However, having severe problems (EQ-5D level 3) in any of the five dimensions assigned could plausibly imply one to appreciate the overall quality of life more than one otherwise would. This argument is related to the above discussion on coping. A lower R-squared for the value set based on the ICUROS data than for the original value set is reasonable, as the sample for estimation of the original algorithm was notably larger.
Adding together the coefficient estimates (levels 2 and 3) to receive each dimension’s total influence in the ICUROS value set estimated with the RE method, self-care appears as the most influential dimension for the resulting QoL utility. The next most influential dimension is anxiety/depression, and the ensuing dimension is pain/discomfort. For the original algorithm, the order of the dimensions’ influence computed in the same way looks quite different; pain/discomfort is in total the most influential, followed by mobility. Absolute differences between the algorithms are largest regarding level 3 of the dimensions. For level 2 of a health state (moderate problems), the actual patient’s view and the view of the speculating, healthy individual do not seem to differ to any appreciable extent; the estimated value set is even implying a lower QoL than the original value set for three out of five dimensions. However, looking at the coefficient estimates for level 3 – the original value set consistently implies a lower QoL than does the estimated value set, which is further illuminated by the three health states given as examples. Investigation of the differences in algorithms further conveys that there are substantial differences between the view of a person hypothesizing upon being in impaired health and the view of a person actually experiencing impaired health. Having determined that the instrument outcomes differ, the question remains of which perception or preferences would be appropriate to apply. In such a matter, different arguments could be reasonable. The healthy population, the foundation of the original EQ-5D set of utility values, consists of the tax payers who finance society’s health-care system and the nursing being performed for the diseased people. On the other hand, only the individual patient has actually experienced the medical condition in question, and only the individual patient knows how that typical health state is perceived in real life. Furthermore, the individual patient is also a tax payer and should benefit the same decision rights as anyone else regarding resource allocation for health care. The individual patient could – on average – also be seen as a representative for the healthy people if they later would enter the state of impaired health in real life.

Patients’ comprehension of how to use the instruments for grading their perceived QoL is of crucial importance. If the study persons do not completely understand the intuition behind e.g. direct TTO, then grading performed may not be reflecting the actual preferences of the individuals. Face-to-face interviews, as were performed in the ICUROS study, could possibly imply a higher accuracy of results from QoL estimations, since the probability of
misunderstandings and misuse thereby may be reduced. Patients included may have other, individual, reasons for changes in HRQoL. Even though exclusion of co-morbidities related to fracture events was performed in the ICUROS study, other QoL-affecting factors may play a part. Patients’ cultural perceptions of QoL may depend on home country, although not to an extent distorting the results received in this thesis. The amount of time spent in a health state may also affect the patient’s perception of QoL and hence his or her QoL estimation. When performing studies of patients afflicted with chronic diseases, it could be of importance to be aware of this possible confounder.

Since this study uses direct TTO grading from actual patients to estimate an EQ-5D value set, it could be considered more accurate than the original value set regarding how the medical conditions and their QoL implications actually are perceived by the patients. To satisfy its purposes, QoL must be a reliable parameter for further analyses to be reliable. In this context, it is hence of importance to determine which type of EQ-5D value set to use – one based on evaluation of actually perceived health states, or one based on evaluation of hypothetical and solely described health states. A suggestion for further research would be to generate EQ-5D value sets based on actual patients instead of healthy people’s hypothetical views of different medical conditions.
8 Conclusions

Comparing mean outcomes from the EQ-5D descriptive system and direct Time-Trade-Off grading performed by the same cohort of patients with osteoporotic fractures, showed that they consistently differ. Sorted by fracture type and time point for utility estimation respectively, QoL estimates from direct TTO were continuously higher than estimates received from the EQ-5D descriptive system and its original algorithm. Estimating a value set, linking EQ-5D scores from the patients to the outcome of direct TTO from the same patients, further illustrated the difference in outcomes, and also showed that increased severity of health state evaluated tended to increase the difference between outcomes from the two methods. Estimation of a patient-based value set also illuminated that for light health states, however, QoL utilities based on EQ-5D’s original algorithm may exceed utilities based on the algorithm that was estimated for actual patients within this study.

The difference in mean outcomes from the two instruments comes to practical life when performing cost-utility analyses. As values from direct TTO indicate a higher perceived QoL than do EQ-5D scores based on the original algorithm, using the former reduces the cost-utility compared to using the latter. In reality, direct TTO are not frequently used in studies, and therefore the practical implications of the difference in outcomes illustrated are not crucial regarding cost-utility evaluations in general. However, assuming that both techniques were in use collaterally, the comparison highlights an important difference in implications depending on technique used. In addition, studies of the burden to society for a specific disease using EQ-5D values would imply increased burden to society compared to when supporting such an analysis with corresponding values from direct TTO. Also the relative burden to society between diseases could change, since the severity of disease is accentuated with usage of EQ-5D values. This thesis however points out an important lesson to learn – estimated quality of life and results from further performed analyses depend on the choice of method used to assess QoL and may differ significantly depending on method.
Acknowledgements

I want to express my gratitude to Fredrik Borgström for his mentorship and the opportunity to access data from the ICUROS study. I am thankful to my supervisor Magnus Johannesson for contributing with his knowledge within the field of health economics. I also want to thank Oskar Ström and Erik Landfeldt who have provided me with knowledgeable comments and recommendations. Last but by no means least; I want to thank all the investigators and health personnel who have enabled the performance of this study.

The ICUROS study was endorsed and partially funded by the International Osteoporosis Foundation (IOF). Research grants were also received from Amgen, Eli Lilly, Medtronic (Kyphon), Novartis, Sanofi-Aventis, Servier and Pfizer.
A EQ-5D Descriptive System

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility
I have no problems in walking about
I have some problems in walking about
I am confined to bed

Self-Care
I have no problems with self-care
I have some problems washing or dressing myself
I am unable to wash or dress myself

Usual Activities (e.g. work, study, housework, family or leisure activities)
I have no problems with performing my usual activities
I have some problems with performing my usual activities
I am unable to perform my usual activities

Pain/Discomfort
I have no pain or discomfort
I have moderate pain or discomfort
I have extreme pain or discomfort

Anxiety/Depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed

© 1990 EuroQol Group EQ-5D™ is a trade mark of the EuroQol Group
B.1 TTO question used for the ICUROS study – at first contact

Time trade off questionnaire

Imagine that your remaining life expectancy is 10 years with your current health status and that you have a choice between two alternatives: Either you will stay in the current health state for 10 years and then die, or you will have full health, but then you will have to give up some years of life. (Please note that this choice is purely hypothetical).

You will thus live for 10 years with current health, or for a shorter period of time in full health. Please write down the number of years in full health that you think is of equal value to 10 years in your current health state. The time scale is there as a visual aid.

Scenario 1. Your health before fracture

Please consider how your health was before the fracture and answer the question in the box below.

Living with health before fracture for 10 years

is equivalent to living in full health for ______ years.

Years
Scenario 2. Your current health

Please relate to your current health and answer the question in the box below.

Living with current health for 10 years

is equivalent to living in full health for _____ years.
B.2 TTO question used for the ICUROS study – at subsequent contacts

Time trade off questionnaire

Imagine that your remaining life expectancy is 10 years with your current health status and
that you have a choice between two alternatives: Either you will stay in the current health
state for 10 years and then die, or you will have full health, but then you will have to give up
some years of life. (Please note that this choice is purely hypothetical).

You will thus live for 10 years with current health, or for a shorter period of time in full
health. Please write down the number of years in full health that you think is of equal value
to 10 years in your current health state. The time scale is there as a visual aid.

Please consider how your current health is and answer the question in the box below.

<table>
<thead>
<tr>
<th>Living with current health for 10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>is equivalent to living in full health for ______ years.</td>
</tr>
</tbody>
</table>
C Inclusion and Exclusion criteria for the ICUROS study

Following criteria were the basis for inclusion and exclusion for the ICUROS study:

- Patients must be aged \( \geq 50 \) years.
- Fracture must be a low-energy fracture.
- Patients are included according to their main diagnosis, patients however excluded if having multiple fractures. Also, patients with fractures caused by co-morbidity, e.g. cancer, should be subject to exclusion.
- Patients are excluded if receiving a new osteoporosis-related fracture during the study period.
- Patients included only if they lived in own housing prior to the event of fracture. Patients also need to be judged as capable of answering the formulary to be included.
- Vertebral fractures should be confirmed by X-ray examination. Fracture of this type does not have to be new, but it however needs to be the first time it is subject for the patient to seek medical attention.
- First interview must be made no more than two weeks after the date of fracture (which is the same as date of first contact with health personnel) for patients to be included.

The physician at each study centre made the decisions which patients to include and exclude based on the above criteria, except for the cases where patients were not judged as capable of answering the formulary. Exclusion of these patients could also be made by the nurse performing the interview.
References

1 Bell DE, Farquhar PH. PERSPECTIVES ON UTILITY THEORY. Operations Research 1986; 34: 179-83


3 Burstrom K, Johannesson M, Diderichsen F. A comparison of individual and social time trade-off values for health states in the general population. Health Policy 2006; 76: 359-70


6 Cullis JG, Jones P. Public finance and public choice. 1998;


8 Dolan P. The measurement of individual utility and social welfare. J Health Econ 1998; 17: 39-52


11 Drummond MF. Methods for the economic evaluation of health care programmes. 1997;
12 Ekman M. Studies in health economics: modelling and data analysis of costs and survival. 2002;


18 Kobelt G. Health economics: an introduction to economic evaluation. 2002;

19 Lai BM, Tsang SW, Lam CL, Kung AW. Validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO-31) in Chinese. Clin Rheumatol 2010; 29: 965-72

20 Landfeldt E, Strom O, Robbins S, Borgstrom F. Adherence to treatment of primary osteoporosis and its association to fractures-the Swedish Adherence Register Analysis (SARA). Osteoporos Int 2011;


25 SBU , Statens beredning för medicinsk utvärdering. [Osteoporosis--prevention, diagnosis and treatment. A systematic literature review. SBU conclusions and summary]. Lakartidningen 2003; 100: 3590-3595


27 Szende Á , Williams A. Measuring self-reported population health: an international perspective based on EQ-5D. 2007;

28 SZENDE AGOT, DEVLIN NANC, OPPE MARK. EQ-5D Value Sets [electronic resource]: Inventory, Comparative Review and User Guide. 2007;


31 Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res 1972; 7: 118-33

32 Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. J Health Econ 2006; 25: 334-46


35 Wooldridge JM. Introductory econometrics: a modern approach. 2009;

36 Zethraeus N, Johannesson M. A comparison of patient and social tariff values derived from the time trade-off method. Health Econ 1999; 8: 541-45