European Heart Journal – Quality of Care and Clinical Outcomes (2016) **2**, 193–200
doi:10.1093/ehjqcco/qcw002

ORIGINAL ARTICLE

Comparison of EQ-5D and 15D instruments for assessing the health-related quality of life in cardiac surgery patients

Jari Heiskanen[1]*, Anna-Maija Tolppanen[2], Risto P. Roine[3,4], Juha Hartikainen[5,6], Mikko Hippeläinen[5,7], Heikki Miettinen[5,8], and Janne Martikainen[9]

[1]Kuopio University Hospital and Research Centre for Comparative Effectiveness and Patient Safety (RECEPS), School of Pharmacy, University of Eastern Finland, Kuopio Campus, Yliopistonranta 1 C, PO Box 1627, Kuopio FI-70211, Finland; [2]Research Centre for Comparative Effectiveness and Patient Safety (RECEPS), School of Pharmacy, University of Eastern Finland, Kuopio, Finland; [3]Research Centre for Comparative Effectiveness and Patient Safety (RECEPS), Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland; [4]Kuopio University Hospital and Helsinki University Hospital, Kuopio, Finland; [5]School of Medicine, University of Eastern Finland, Kuopio, Finland; [6]Heart Center, Kuopio University Hospital, Kuopio, Finland; [7]Heart Surgery, Heart Center, Kuopio University Hospital, Kuopio, Finland; [8]Department of Cardiology, Heart Center, Kuopio University Hospital, Kuopio, Finland; and [9]School of Pharmacy, Pharmacoeconomics and Outcomes Research Unit (PHORU), University of Eastern Finland, Kuopio, Finland

Received 16 November 2015; revised 22 December 2015; online publish-ahead-of-print 5 January 2016

See page 149 for the editorial comment on this article (doi:10.1093/ehjqcco/qcw015)

Aims Patient-centred outcomes can be measured with different instruments. We compared the performance of two health-related quality-of-life (HRQoL) measures, EQ-5D and 15D, in patients undergoing elective coronary artery bypass grafting (CABG).

Methods and results Patients who were admitted for elective CABG in Kuopio University Hospital Finland in 2012–14 and had completed both instruments concurrently as part of the admission process ($n = 182$). Follow-up was conducted by postal survey 12 months after the CABG operation. The validity, agreement, and responsiveness to change of both instruments were examined. The mean baseline HRQoL index scores obtained by the EQ-5D and the 15D were 0.795 and 0.859, respectively ($P < 0.001$ for difference). The agreement between instruments was poor (Spearman's rho = 0.449; $P < 0.001$). Observed ceiling effects at baseline for the EQ-5D and 15D were 31.9 and 4.4%, respectively. EQ-5D was able to discriminate distinct Canadian Cardiovascular Society groups. During the 1-year follow-up, clinically important improvement was observed in 39.6 and 53.3% of patients with the EQ-5D and the 15D, respectively. However, with the 15D, the number of operated patients required to produce one additional quality-adjusted life year (QALY) was more than twice as high compared with the EQ-5D.

Conclusion EQ-5D and 15D do not appear to be interchangeable when patient-centred outcomes in CABG patients are assessed. The EQ-5D seems to have better discriminative power and known-group validity, whereas the 15D is more sensitive to change over time. These instruments lead to significantly different estimates concerning the number of QALYs gained.

Keywords Coronary disease • Revascularization • Coronary artery bypass graft surgery

Introduction

Coronary artery disease (CAD) is a common and costly disease. In developed countries, it causes approximately one-fourth of all deaths. It can be treated with revascularization either by coronary artery bypass grafting (CABG) or percutaneous coronary intervention, or by conventional pharmacotherapy.[1] CABG is generally preferred for patients with left main CAD or more advanced disease.[2–8]

After a cardiac intervention, outcomes are commonly evaluated in terms of mortality, complications, recurrence of symptoms, or changes in functional capacity, as all of them can be measured relatively easily. However, in the era of patient-centred healthcare, also patient-reported outcomes such as changes in physical, psychological, and social functioning are deemed important. In addition, as survival rates have improved significantly, survival alone is no longer the only goal of treatment and health-related quality of life (HRQoL) also plays an essential role.[9]

* Corresponding author. Tel: +358 50 362 3058, Fax: +358 17 162 131, Email: jari.heiskanen@uef.fi

Currently, there are several preference-based HRQoL measures, such as the SF-6D,[10] HUI3,[11] EQ-5D,[12,13] and the 15D,[14–16] that provide a descriptive health state profile and a utility score of overall HRQoL. The EQ-5D appears to be currently the most frequently used preference-based instrument worldwide,[17] whereas the 15D is a widely used HRQoL instrument in Finnish hospitals. Both of these instruments provide single utility scores that can be applied to the estimation of quality-adjusted life years (QALYs). The QALY concept is widely used to assess the value for money of health technologies and one of its main advantages is that it combines both survival and HRQoL benefits of treatments in a single indicator.

The EQ-5D has been compared against other HRQoL measures,[18,19] but studies comparing the EQ-5D and 15D are less common. To our knowledge, only one previous study has assessed the similarity between the EQ-5D and 15D scales in patients undergoing elective CABG.[20] Moreover, little is known on how the selection of HRQoL instruments is related to clinically meaningful changes in HRQoL and the number of QALYs gained in this same patient group. Therefore, the purpose of the present study was to explore the similarities and differences of the EQ-5D and 15D instruments in measuring patient-centred outcomes (in terms of HRQoL) of elective CABG. The comparison was made based on validity, degree of agreement, responsiveness, percentage of patients with a clinically meaningful change in HRQoL, and the number of QALYs gained during the 1 year of follow-up.

Methods

Study design and setting

The present study was an observational, methodological study conducted as a part of routine clinical practice. Study participants were recruited from the Heart Center of the Kuopio University Hospital, Kuopio, Finland. The recruitment of patients took place as part of routine clinical practice from 2012 to 2014. Patients admitted for CABG operation were asked to complete both the EQ-5D and 15D instruments concurrently as part of the preoperative hospital admission process. Follow-up was conducted by postal survey 12 months after the operation. To enable comparison of the responsiveness of the two instruments, only respondents fully completing both questionnaires at baseline and at 12 months were included in this study. The demographic and preoperative characteristics of patients were extracted from electronic patient records and linked with the outcome measurements by applying personal identification numbers. All personal identifiers were removed from the final dataset. The study was approved by the ethics committee of the Kuopio University Hospital.

Health-related quality-of-life instruments

The EQ-5D measures mobility, self-care, usual activities, pain/discomfort, and anxiety/depression and is a widely used, self-administrated, generic, five-dimension questionnaire for assessing HRQoL. Each dimension includes three ordinal categories of severity corresponding to no, moderate, or severe problems. Combining one level from each dimension defines 243 different health states ranging from 11111 (full health) to 33333 (worst health). These health states are converted into a single index score representing health utilities (−0.59 to 1.00), using valuations elicited from a sample of the general public. In the present study, the time-trade off (TTO) valuation based on samples of the United Kingdom (UK) general public was applied.[12,13]

The 15D instrument is also a generic, self-administrated questionnaire for measuring HRQoL. It consists of 15 dimensions (mobility, vision, hearing, breathing, sleeping, eating, speech, excretion, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity) with five ordinal levels. The single index score of the 15D instrument ranges from 0 to 1. The 15D instrument can generate over 30 billion different health states, and is generated from a set of utility or preference weights. The valuation system of the 15D used in this study is based on a set of Finnish population-based preferences.[14–16]

Statistical analyses

Baseline demographic data are presented using percentages, means and standard deviations as appropriate. Due to skewed distribution of HRQoL scores, non-parametric tests were applied. Differences between the instruments at baseline were tested by the Wilcoxon signed-rank test. The ceiling effects of the EQ-5D and 15D were assessed by computing the percentage of respondents reporting full health at baseline or at the 12 months. The HRQoL data were reported as means and confidence intervals at baseline and at 12-month follow-up. A non-parametric bootstrap procedure, based on 100 replications, was performed to estimate the mean difference and 95% CI between 12-month changes in EQ-5D and 15D scores. Differences between instruments were tested by the Wilcoxon signed-rank test. In addition, the changes in mean scores were categorized and compared according to minimal important differences (MIDs) reported in the literature for both instruments. The level of statistical significance is defined as $P < 0.05$. All statistical analyses were conducted by STATA 12.0 (Stata Corp. LP, Station, TX, USA) and SPSS 19 (IBM SPSS Statistics).

Construct and discriminant validity

The construct validity of the instruments was assessed by examining Spearman correlation between the estimated utility scores at baseline. Discriminant validity was assessed by the known-group method by investigating whether the EQ-5D and 15D scores are different for predefined distinctive groups. Patients were grouped according to sex, age (<60, 60–74, and ≥75 years), and functional status at baseline. The functional status was defined by the Canadian Cardiovascular Society (CCS) classes for grading angina pectoris.[21] The CCS grading system is analogical with the New York Heart Association grading system.[22,23] Classifications were done preoperatively by heart surgeons. Patients aged ≥75 years and with poorer CCS status were hypothesized to have lower utility scores for these two instruments.

Agreement

Agreement between the utility instruments was evaluated by concordance correlation coefficient (CCC)[24] and the Bland–Altman plot.[25–28] The strength of agreement was considered as poor when CCC <0.9, moderate when CCC 0.90–0.95, and strong when CCC >0.99.[29] In the Bland–Altman plot, the differences between the two utility scores (on the *y*-axis) were plotted against the average values of these utility scores (on the *x*-axis). The deviation of the difference from zero line, which implies total agreement between the instruments, indicates the degree of agreement for each patient on the plot.[25–28]

Responsiveness

To evaluate the responsiveness to change (i.e. ability to detect changes in utility scores over time) of the EQ-5D and 15D, the changes from baseline were estimated. In addition, the change in mean scores were categorized and compared according to MIDs reported in the literature

Table 1 Baseline characteristics of the patients

Variables	Study population (n = 182)	Total population (n = 1018)
Male, n (%)	148 (81.3)	815 (80.1)
Age, mean (SD)	65.8 (9.6)	67.37 (8.5)
Height, mean (SD)	171.9 (11.2)	170.76 (8.4)
Weight, mean (SD)	83.2 (16.6)	82.00 (15.0)
15D baseline mean index (SD, n)	0.859 (0.099) ($n = 182$)	0.852 (0.092) ($n = 773$)
EQ-5D baseline mean index (SD, n)	0.795 (0.207) ($n = 182$)	0.791 (0.195) ($n = 556$)
Canadian Cardiovascular Society class, n (%)		
1	10 (5.5)	17 (1.7)
2	58 (31.9)	228 (22.4)
3	75 (41.2)	400 (39.3)
4	39 (21.4)	368 (36.1)
Missing	0 (0.0)	5 (0.5)
Ejection fraction, n (%)		
≤50 Reduced	58 (31.9)	213 (21.0)
>50 Normal	122 (67.0)	798 (78.4)
Data missing	2 (1.1)	7 (0.7)
Left main stenosis, n (%)		
No	152 (83.5)	620 (60.9)
Yes (over 50%)	30 (16.5)	398 (39.1)
Comorbidities, n (%)[a]		
Yes	64 (35.2)	406 (39.9)
No	118 (64.8)	600 (58.9)
Missing	0 (0.0)	12 (1.2)

[a]Cerebral haemorrhage, central nervous system disease, kidney disease, lung disease, diabetes, potential heart failure, and other disease.

for both instruments. The applied MID limits were 0.074 for the EQ-5D[30,31] and 0.015 for the 15D,[32] respectively.

Differences in the responsiveness between the instruments were compared across patient groups with varying levels of disease severity to reveal whether the magnitude of instrument discrepancy is related to the disease severity. The patients were divided into groups according to their ejection fraction (>50%; normal, ≤50%; mild/moderate reduction, <30%; severe reduction), the diagnosis of left main stenosis, the presence of other comorbidities (i.e. cerebral haemorrhage, central nervous system disease, kidney disease, lung disease, diabetes, potential heart failure, and other disease), age (cut-off 75 years), sex, disease severity (CCS class ≥3), and body mass index (BMI; cut-off 25). Differences between the predefined groups were tested using Fisher's exact test.

Effect of instrument on the number of minimal important differences and quality-adjusted life years gained

The inverted values of the proportion of patients reaching the MID at 12 months were applied to describe the number of required persons to reach one positive MID by both instruments. To demonstrate the impact of differences in the EQ-5D and 15D utility scores on the number of QALYs gained during the 1-year follow-up, the number needed per QALY gained (NNQ) approach was applied.[33] The NNQ is a group-level estimate for the number of patients that must be operated in order to gain 1 QALY and thus, it provides a simple and practical metric to compare the results obtained by two different instruments. The NNQ was estimated as an inverted value of the utility gain measured by the EQ-5D and 15D instruments (i.e. 1/mean change from baseline in utility score at 12 months).

Results

Between years 2012 and 2014, a total of 1018 CABG operations were conducted in the Kuopio University Hospital. The final study sample included 182 patients with HRQoL data measured with both instruments at baseline and at the 12-month follow-up (17.8% of eligible patients). The demographic and preoperative characteristics of the total population and the present study sample are described in *Table 1*. The study participants were slightly younger than the total population on average. In addition, the proportions of patients in different CCS classes and the prevalence of other clinical conditions indicated that the sample of study participants represented a patient group with slightly less severe disease states when compared with the total population. However, no significant differences in HRQoL index values between the total and study populations were observed (*Table 1*).

At baseline, the mean (95% CI) EQ-5D and 15D scores were 0.795 (0.765–0.826) and 0.859 (0.845–0.874), respectively. Thus, the 15D produced significantly higher mean baseline scores than the EQ-5D, with the mean difference (15D minus EQ-5D) (95% CI) being 0.064 (0.040–0.088) ($P < 0.001$ for difference). Differences between the instruments remained across the majority of

Table 2 Baseline scores for the EQ-5D and 15D stratified by the baseline characteristics

Variable	n (%)	15D baseline utility score	EQ-5D baseline utility score	Mean utility score difference between 15D and EQ-5D at baseline[a]	CCC between 15D and EQ-5D at baseline[b]	Spearman's correlations[c]
Sex						
Men	148 (81.3)	0.866	0.807	0.059 ($P < 0.001$)	0.517 ($P < 0.001$)	0.593 ($P < 0.001$)
Women	34 (18.7)	0.832	0.743	0.089 ($P = 0.081$)	0.250 ($P = 0.010$)	0.651 ($P < 0.001$)
P-value*		*0.056*	*0.104*			
Age (years)						
<60	43 (23.6)	0.886	0.843	0.042 ($P = 0.128$)	0.436 ($P < 0.001$)	0.613 ($P < 0.001$)
60–74.9	103 (56.6)	0.853	0.776	0.077 ($P < 0.001$)	0.461 ($P < 0.001$)	0.690 ($P < 0.001$)
≥75	36 (19.8)	0.848	0.794	0.053 ($P = 0.019$)	0.310 ($P = 0.020$)	0.439 ($P = 0.007$)
P-value[†]		*0.190*	*0.193*			
CCS						
1	10 (5.5)	0.903	0.908	−0.005 ($P = 0.838$)	0.385 ($P = 0.134$)	0.555 ($P = 0.096$)
2	58 (31.9)	0.873	0.843	0.031 ($P = 0.048$)	0.543 ($P < 0.001$)	0.582 ($P < 0.001$)
3	75 (41.2)	0.851	0.780	0.071 ($P < 0.001$)	*0.488* ($P < 0.001$)	*0.618* ($P < 0.001$)
4	39 (21.4)	0.846	0.729	0.119 ($P = 0.008$)	0.357 ($P < 0.001$)	0.618 ($P < 0.001$)
P-value[†]		*0.261*	*0.039*			

CCC, concordance correlation coefficient; CCS, Canadian Cardiovascular Society.
[a]According to the Wilcoxon signed-rank test.
[b]Concordance correlation coefficient (CCC), rho.
[c]Spearman's and Kendall's correlations, Spearman's rho.
*According to the two-sample Wilcoxon rank-sum (Mann–Whitney) test.
[†]According to the Kruskal–Wallis equality-of-populations rank test.

the defined subgroups (*Table 2*). Neither the 15D, nor the EQ-5D, exhibited a significant floor effect in this study. However, at baseline, the ceiling effects for the EQ-5D and the 15D were 31.9 and 4.4%, respectively.

Construct and discriminant validity

The EQ-5D and 15D produced significantly different baseline utility scores across different subgroups at baseline. There was a tendency for lower utility scores among women and older individuals. Only the EQ-5D was able to discriminate the different CCS classes ($P = 0.039$), indicating disease severity at baseline (*Table 2*).

Agreement between instruments

The CCCs demonstrated poor agreement between the instruments (*Table 2*). *Figure 1* graphically presents the discrepancy between the EQ-5D and 15D instruments at baseline. According to the Bland–Altman plot, the limits of agreement (LOAs) were −0.454 to 0.278 for a mean difference (95% CI) of 0.064 (0.040–0.088), which is equal to an expected between-measure variation of 0.732 (i.e. a range between −0.454 and 0.278) for any pair of future baseline observation (*Figure 1*). Thus, also the LOA indicated large differences between these two instruments for individual subjects.

Responsiveness to change

The observed mean changes (95% CI) from baseline to the 12-month follow-up were 0.053 (0.017–0.088) and 0.024 (0.009–0.038) for the EQ-5D and 15D, respectively ($P = 0.024$ for difference). The correlation between 1-year change in EQ-5D and 15D was 0.476, $P < 0.001$ (*Figure 2*). When the observed changes were stratified according to the MID threshold values, the EQ-5D indicated improvement in 39.6% (95% CI 32.4–46.7%) of the patients in contrast to the 53.3% (46.0–60.6%) by the 15D.

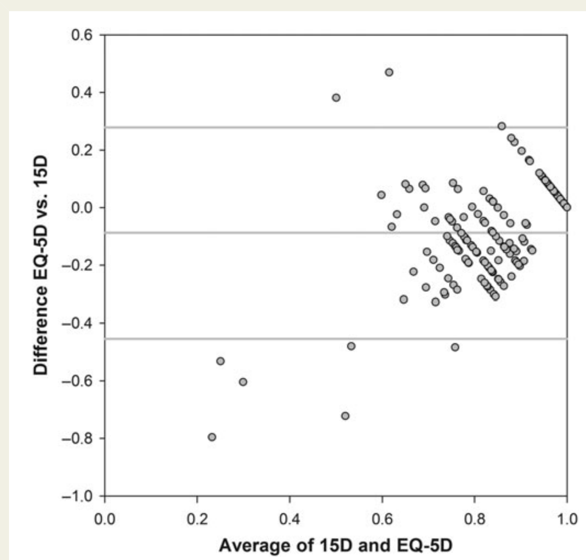

Figure 1 Bland–Altman plot for EQ-5D and 15D scores at baseline.

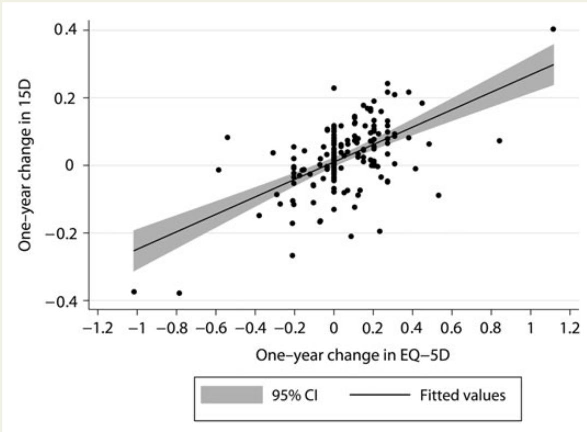

Figure 2 Correlation between 1-year change in 15D and EQ-5D.

Furthermore, the EQ-5D indicated no clinically meaningful change in 45.6% (38.3–52.9%) of the patients as opposed to 22.0% (15.9–28.1%) by the 15D. A clinically important deterioration was reported by 14.8% (9.6–20.0%) of the patients with the EQ-5D and by 24.7% (18.4–31.1%) with the 15D, respectively. The proportions of changes stratified according to the MID values were significantly different between the instruments ($P < 0.001$ for difference).

Significant differences between instruments were systematically observed when the stratified MIDs were grouped by severity of illness like left main stenosis, comorbidities, age group (cut-off >75 years), lower than 50% ejection fraction, CCS classes 3 and 4, BMI (cut-off >25), or sex (*Table 3*). Both instruments indicated that poorer preoperative status (i.e. left main stenosis or poor ejection fraction) was associated with a higher percentage of improved patients (i.e. patients in whom the change in HRQoL exceed the MID) when compared with the whole study sample. More interestingly, both instruments also indicated that among a subgroup of patients aged over 75 years, the relative proportion of improved patients was lower than in the whole study sample. Although we did not observe any differences in HRQoL change between men and women (data not shown), the instruments performed similarly in women but differently in men.

Effect of instrument on the number of minimal important differences and quality-adjusted life years gained

The number of patients needed to be treated to produce one patient with a positive MID at 12 months was rather similar for both instruments with no significant differences (*Table 4*). As summarized in *Table 4*, on average, the difference between the numbers of patients needed to treat was only one. However, the NNQ estimates demonstrated much wider and significant disagreement between the instruments. With the 15D, the number of operated patients required to produce one additional QALY was more than twice as high compared with the EQ-5D.

Discussion

Our findings show that the reported pre- and postoperative utilities, as well as proportions of significant quality of life and QALY gains from a CABG operation, are heavily dependent on the measure used to elicit them. The EQ-5D produced significantly lower preoperative utility scores than the 15D instrument for patients undergoing CABG operation. Among these CABG patients, the EQ-5D showed better discriminative power as it was able to distinguish the preoperative CCS classes better than the 15D instrument. The degree of disagreement between the instruments showed that the variation between the measures is too great for them to be considered interchangeable. There are many possible reasons for these differences in HRQoL measured with EQ-5D and 15D, including the dimensions, valuation sets, and applied scales of different instruments. Our findings are in line with a previous study reporting only a moderate agreement between the instruments,[34] with even lower utility scores among CAD patients (0.684 with the EQ-5D and 0.821 with the 15D) than observed in our study.[35]

As expected, larger 12-month changes from baseline were observed with the EQ-5D than the 15D due to the differences in their valuation processes and different theoretical scales of measurement. Actually, the magnitude of the overall utility gain was twice as large when measured with the EQ-5D compared with the 15D, even if the EQ-5D produced lower values in general. Thus, when the 15D results were used as the basis for calculating the NNQ roughly twice as many operated patients were needed to produce one QALY compared with those obtained with the EQ-5D. In addition, this difference has also important implications for cost–utility analyses: similar to previous research, our study shows that the choice of instrument can have a crucial effect on the results of cost–utility analysis.[34] Thus, vigilance is warranted when deciding which HRQoL instrument to use.[35,36]

In daily clinical practice, the proportion of patients reaching the positive MID may act as a more practical and patient-centred outcome for monitoring clinical success than the cumulative numbers of QALYs gained. However, according to our findings, these two metrics produce conflicting results: the EQ-5D instrument leads to a larger absolute improvement in the utility score and a higher number of QALYs gained, but to a smaller proportion of patients reaching the positive MID than the 15D. Thus, the interpretation is different when the observed changes in the mean utility scores are related to the reported MID values of the instruments, i.e. the 15D seems to be more sensitive to change and indicates improvement in health utility more often than the EQ-5D does. Furthermore, our subgroup analyses revealed that if the health status is better than average, the EQ-5D cannot properly differentiate changes in the reported health state because of its ceiling effect as reported previously.[37] Thus, in poor health states, the EQ-5D can indicate very low values, even below zero, which is equivalent to a health state poorer than death. The EQ-5D responds easily to poor health states, but at the same time overreacts to good health by producing easily full index scores. These properties of the EQ-5D have also been observed earlier in comparison with the SF-6D.[38]

The results of the study need to be interpreted in light of some limitations. First, we only included the respondents who had fully completed both questionnaires at baseline and 12-month follow-up.

Table 3 Twelve-month changes from baseline measured by the 15D and the EQ-5D in different subgroups

	15D (MID 0.015)[a], *n* (%)	EQ-5D (MID 0.074)[a], *n* (%)	*P*-values[†]
All patients			
Negative MID change	45 (24.7)	27 (14.8)	<0.001
No MID changes	40 (22.0)	83 (45.6)	
Positive MID change	97 (53.3)	72 (39.6)	
Total, *n*	182	182	
Age group over 75 years			
Negative MID change	15 (41.7)	11 (30.6)	0.002
No MID changes	8 (22.2)	15 (41.7)	
Positive MID change	13 (36.1)	10 (27.8)	
Total, *n*	36	36	
Sex (female)			
Negative MID change	4 (11.8)	6 (17.7)	0.07
No MID changes	11 (32.3)	16 (47.0)	
Positive MID change	19 (55.9)	12 (35.3)	
Total, *n*	34	34	
Sex (male)			
Negative MID change	41 (27.7)	21 (14.2)	<0.001
No MID changes	29 (19.6)	67 (45.3)	
Positive MID change	78 (52.7)	60 (40.5)	
Total, *n*	148	148	
BMI >25			
Negative MID change	33 (25.4)	19 (14.6)	<0.001
No MID changes	29 (22.3)	56 (43.1)	
Positive MID change	68 (52.3)	55 (42.3)	
Total, *n*	130	130	
Comorbidities			
Negative MID change	18 (28.1)	8 (12.5)	0.001
No MID changes	13 (20.3)	28 (43.8)	
Positive MID change	33 (51.6)	28 (43.8)	
Total, *n*	64	64	
Ejection fraction reduced			
Negative MID change	11 (19.0)	5 (8.6)	0.006
No MID changes	9 (15.5)	25 (43.1)	
Positive MID change	38 (65.5)	28 (48.3)	
Total, *n*	58	58	
Left main stenosis population			
Negative MID change	6 (20.0)	2 (6.7)	0.012
No MID changes	5 (16.7)	12 (40.0)	
Positive MID change	19 (63.3)	16 (53.3)	
Total, *n*	30	30	
CCS class 3 or 4			
Negative MID change	30 (26.3)	15 (13.2)	0.001
No MID changes	23 (20.2)	50 (43.9)	
Positive MID change	61 (53.5)	49 (43.0)	
Total, *n*	114	114	

MID, minimal important difference; BMI, body mass index; CCS, Canadian Cardiovascular Society.
[a]Based on literature.[27–29]
[†]Fisher's exact test.

Table 4 Number of patients needed to be operated for one additional clinically important change in health-related quality of life or and one QALY to be gained

Utility score instrument	Number needed to treat for MID[a]	95% CI	NNQ[b]	95% CI
EQ-5D	3	2.14–3.09	19	11.82–48.46
15D	2	1.65–2.17	43	26.39–108.50

NNQ, number needed per QALY gained; QALY, quality-adjusted life years; MID, minimal important difference.
[a]Number needed to be treated to reach one positive MID.
[b]The number of patients needed to be operated for one additional QALY to be gained (NNQ).

This might increase the risk of selection bias and limit the generalizability of our results. However, the characteristics of those included in the study sample ($n = 182$) were comparable to the eligible population (i.e. those undergoing CABG), suggesting that the results are generalizable to the target population (i.e. those undergoing CABG). One potential limitation is also the use of the scoring system of the EQ-5D which is based on the UK TTO system. However, there is no local TTO algorithm available for the EQ-5D in Finland, and therefore UK TTO valuations have also been applied in a previous Finnish study.[35] One limitation is also that we did not apply the newer EQ-5D-5L version in the present study, even if it has been shown to be promising compared with the EQ-5D-3L version in terms of a lower ceiling effect, better discriminatory power, and known-groups validity.[39] However, currently, the country-specific value sets for the new EQ-5D-5L are lacking for many countries. Therefore, the use of the newer EQ-5D version is limited until country-specific value sets are developed.

As it may be too optimistic to hope that healthcare providers will reach a consensus regarding which instruments to use for measurement of patient-centred outcomes in terms of HRQoL, it is not unreasonable to expect that at least the applied HRQoL instruments and methods are clearly and transparently stated when patient-centred outcome studies are reported. Otherwise, there will be a lot of data that are non-comparable and unusable for the comparison of different hospitals. Even in the case of one hospital and one disease, it is necessary to raise awareness of the different performance of HRQoL instruments to ensure rational decision-making, although the choice of the HRQoL instrument may well be based on study objectives as previously suggested.[40]

Conclusion

In CABG patients, the EQ-5D seems to have better discriminative power and known-group validity, whereas the 15D is more sensitive to change over time. The use of these instruments in the estimation of QALYs gained leads to significantly different estimates. Overall, the EQ-5D and 15D do not appear to be interchangeable.

Acknowledgements

The authors thank for clinical study nurse, RN, Lari Kujanen (Kuopio University Hospital Heart Center) and project coordinator, MSc (Health Econ.), RN, Ninna Mäkirinne-Kallio for their help in the collection of clinical and quality-of-life data.

Funding

R.P.R. acknowledges financial support by European Structural Funds, European Social Fund by the Regional Council of North-Savo, and State Research Funding (VTR). The funders had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript. The views expressed in this paper are those of the authors and not necessarily those of any funding body or others whose support is acknowledged.

Conflict of interest: J.M. is a partner of ESiOR Oy, which carries out health economic and outcome research studies for pharmaceutical and food companies. J.He. drafted the first version on the manuscript, performed statistical analyses, acts as a guarantor (had full access to data), revised the draft version, and accepted the final manuscript. A.-M.T. planned the research project, revised the draft version, and accepted the final version. R.P.R. revised the draft version and accepted the final version. J.Ha. planned the research project, revised the draft version, and accepted the final version. M.H. revised the draft version and accepted the final version. H.M. revised the draft version and accepted the final version. J.M. planned the research project, supervised statistical analyses, revised the draft version, and accepted the final version.

References

1. Task Force on Myocardial Revascularization of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS), European Association for Percutaneous Cardiovascular Interventions (EAPCI), Wijns W, Kolh P, Danchin N, Di Mario C, Falk V, Folliguet T, Garg S, Huber K, James S, Knuuti J, Lopez-Sendon J, Marco J, Menicanti L, Ostojic M, Piepoli MF, Pirlet C, Pomar JL, Reifart N, Ribichini FL, Schalij MJ, Sergeant P, Serruys PW, Silber S, Sousa Uva M, Taggart D. Guidelines on myocardial revascularization. *Eur Heart J* 2010;**31**:2501–2555.
2. Hlatky MA, Boothroyd DB, Baker L, Kazi DS, Solomon MD, Chang TI, Shilane D, Go AS. Comparative effectiveness of multivessel coronary bypass surgery and multivessel percutaneous coronary intervention: a cohort study. *Ann Intern Med* 2013; **158**:727–734.
3. Mohr FW, Morice MC, Kappetein AP, Feldman TE, Stahle E, Colombo A, Mack MJ, Holmes DR Jr, Morel MA, Van Dyck N, Houle VM, Dawkins KD, Serruys PW. Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *Lancet* 2013;**381**:629–638.
4. Sa MP, Ferraz PE, Escobar RR, Nunes EO, Soares AM, de Araujo e Sa FB, Vasconcelos FP, Lima RC. Five-year outcomes following PCI with DES versus CABG for unprotected LM coronary lesions: meta-analysis and meta-regression of 2914 patients. *Rev Bras Cir Cardiovasc* 2013;**28**:83–92.
5. Weintraub WS, Grau-Sepulveda MV, Weiss JM, O'Brien SM, Peterson ED, Kolm P, Zhang Z, Klein LW, Shaw RE, McKay C, Ritzenthaler LL, Popma JJ, Messenger JC, Shahian DM, Grover FL, Mayer JE, Shewan CM, Garratt KN, Moussa ID, Dangas GD, Edwards FH. Comparative effectiveness of revascularization strategies. *N Engl J Med* 2012;**366**:1467–1476.
6. 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**:94–99.

7. Serruys PW, Unger F, Sousa JE, Jatene A, Bonnier HJ, Schonberger JP, Buller N, Bonser R, van den Brand MJ, van Herwerden LA, Morel MA, van Hout BA, Arterial Revascularization Therapies Study Group. Comparison of coronary-artery bypass surgery and stenting for the treatment of multivessel disease. *N Engl J Med* 2001; **344**:1117–1124.
8. Rogers CA, Pike K, Campbell H, Reeves BC, Angelini GD, Gray A, Altman DG, Miller H, Wells S, Taggart DP, CRISP investigators. Coronary artery bypass grafting in high-RISk patients randomised to off- or on-Pump surgery: a randomised controlled trial (the CRISP trial). *Health Technol Assess* 2014;**18**(v–xx):1–157.
9. Rumsfeld JS, Alexander KP, Goff DC Jr, Graham MM, Ho PM, Masoudi FA, Moser DK, Roger VL, Slaughter MS, Smolderen KG, Spertus JA, Sullivan MD, Treat-Jacobson D, Zerwic JJ, American Heart Association Council on Quality of Care and Outcomes Research, Council on Cardiovascular and Stroke Nursing, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, and Stroke Council. Cardiovascular health: the importance of measuring patient-reported health status: a scientific statement from the American Heart Association. *Circulation* 2013;**127**:2233–2249.
10. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;**21**:271–292.
11. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health utilities index. *Pharmacoeconomics* 1995;**7**:503–520.
12. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72.
13. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208.
14. Sintonen H. The 15D-measure of health-related quality of life. I. Reliability, validity and sensitivity of its health state descriptive system. 1994; Working Paper 41.
15. Sintonen H. The 15D-measure of health-related quality of life. II. Feasibility, reliability and validity of its valuation system. 1995; Working Paper 42.
16. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med* 2001;**33**:328–336.
17. Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen 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**:235–241.
18. De Smedt D, Clays E, Doyle F, Kotseva K, Prugger C, Pajak A, Jennings C, Wood D, De Bacquer D, EUROASPIRE Study Group. Validity and reliability of three commonly used quality of life measures in a large European population of coronary heart disease patients. *Int J Cardiol* 2013;**167**:2294–2299.
19. van Stel HF, Buskens E. Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health Qual Life Outcomes* 2006;**4**:20.
20. Kattainen E, Sintonen H, Kettunen R, Merilainen 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**: 172–179.
21. Campeau L. Letter: grading of angina pectoris. *Circulation* 1976;**54**:522–523.
22. The Criteria Committee of the New York Heart Association. *Nomenclature and criteria for diagnosis of diseases of the heart and blood vessels*, 6th ed. Boston: Little Brown; 1964.
23. The Criteria Committee of the New York Heart Association, ed. *Nomenclature and criteria for the diagnosis of diseases of the heart and great vessels*, 9th ed. Boston: Little Brown & Co; 1994.
24. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;**45**:255–268.
25. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983;**32**:307–317.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–310.
27. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;**8**:135–160.
28. Bland JM, Altman DG. Agreed statistics: measurement method comparison. *Anesthesiology* 2012;**116**:182–185.
29. McBride GB. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. 2005; HAM2005–062.
30. 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**:1523–1532.
31. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes* 2007;**5**:70.
32. Alanne S, Roine RP, Rasanen P, Vainiola T, Sintonen H. Estimating the minimum important change in the 15D scores. *Qual Life Res* 2015;**24**:599–606.
33. Gulfe A, Kristensen LE, Saxne T, Jacobsson LT, Petersson IF, Geborek P. Utility-based outcomes made easy: the number needed per quality-adjusted life year gained. An observational cohort study of tumor necrosis factor blockade in inflammatory arthritis from Southern Sweden. *Arthritis Care Res (Hoboken)* 2010;**62**: 1399–1406.
34. Vainiola T, Pettila V, Roine RP, Rasanen 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**:2090–2093.
35. Saarni SI, Harkanen T, Sintonen H, Suvisaari J, Koskinen S, Aromaa A, Lonnqvist 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**: 1403–1414.
36. 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**:485–495.
37. Macran S, Weatherly H, Kind P. Measuring population health: a comparison of three generic health status measures. *Med Care* 2003;**41**:218–231.
38. Kontodimopoulos N, Argiriou M, Theakos N, Niakas D. The impact of disease severity on EQ-5D and SF-6D utility discrepancies in chronic heart failure. *Eur J Health Econ* 2011;**12**:383–391.
39. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res* 2013; **22**:1717–1727.
40. Linde L, Sorensen J, Ostergaard M, Horslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. *J Rheumatol* 2008;**35**:1528–1537.