Paul McCrone, Anita Patel, Martin Knapp, Aart Schene, Maarten Koeter, Francesco Amaddeo, Mirella Ruggeri, Anne Giessler, Bernd Puschner, Graham Thornicroft

A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia

Article (Accepted version)
(Refereed)

This version available at: http://eprints.lse.ac.uk/25437/
Available in LSE Research Online: August 2012

http://eprints.lse.ac.uk

A Comparison of SF-6D and EQ-5D Utility Scores in a Study of Patients with Schizophrenia

[Short title: Comparison of SF-6D and EQ-5D]

Paul McCrone, Health Service and Population Research Department, Institute of Psychiatry, King's College London

Anita Patel, Health Service and Population Research Department, Institute of Psychiatry, King's College London

Martin Knapp, Health Service and Population Research Department, Institute of Psychiatry, King's College London & Personal Social Services Research Unit, London School of Economics and Political Science

Aart Schene, Department of Psychiatry, Academic Medical Centre, University of Amsterdam

Maarten Koeter, Department of Psychiatry, Academic Medical Centre, University of Amsterdam

Francesco Amaddeo, Section of Psychiatry and Clinical Psychology, University of Verona

Mirella Ruggeri, Section of Psychiatry and Clinical Psychology, University of Verona

Anne Giessler, Clinic of Psychotherapy and Psychosomatic Medicine, University of Leipzig

Bernd Puschner, Department of Psychiatry II, Ulm University

Graham Thornicroft, Health Service and Population Research Department, Institute of Psychiatry, King's College London

For correspondence: Paul McCrone, PO24 Health Services and Population Research Department, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK. Telephone: +44 20 7848 0874. E-mail: p.mccrone@iop.kcl.ac.uk

Abstract

Background

Economic evaluations of healthcare interventions increasingly measure outcomes using quality-adjusted life years (QALYs). The SF-6D and the EQ-5D are alternative ways of generating utility scores for use in QALY estimations, but it is unclear which is most sensitive to change in psychiatric symptom severity. There are also limited data on the sensitivity of these measures to changes in existing clinical indicators in long-term mental health conditions like schizophrenia.

Aims of the study

To: (i) assess the relationship between SF-6D and EQ-5D utility scores for patients with schizophrenia at two points in time, (ii) assess the relationship in the change scores of these two measures, (iii) measure the sensitivity of these measures to changes in an established measure of symptomatology.

Methods

Patients with schizophrenia were recruited and the SF-36 and EQ-5D were administered at baseline and one-year follow-up and utility scores were computed and compared.. Standardised response mean (SMR) scores were calculated for the SF-6D and EQ-5D and compared for patients who improved or deteriorated by at least 25% on the Brief Psychiatric Rating Scale.

Results

EQ-5D ratings were available for 394 patients at baseline, 368 at follow-up and 358 at both time points. The respective figures for the SF-6D were 383, 367 and 345. Mean utility scores were very similar at baseline (EQ-5D 0.68, SF-6D 0.67) and follow-up (EQ-5D 0.71, SF-6D

0.68). Median scores were markedly higher for the EQ-5D (0.76 v 0.66 at baseline, 0.80 v 0.68 at follow-up). The SF-6D scores followed a normal distribution whilst the EQ-5D scores were negatively skewed with a clustering at 1.00. There were few differences in sensitivity to change between the EQ-5D and SF-6D.

Discussion

From an analytical perspective the SF-6D has advantages over the EQ-5D due to its normal distribution and lack of ceiling effect. However, both measures produce similar mean utility scores. Overall the SF-6D appears more suitable as a measure if utility in this patient group.

Implications for health policies

Decisions made on the basis of cost-effectiveness results need to consider the method by which QALYs have been calculated.

Implications for further research

Further comparisons of the EQ-5D and SF-6D are required.

Introduction

Deciding which outcome measures are to be used in economic evaluations is important for the purposes of making policy decisions, and different measures can produce very different findings [1]. As well as condition specific measures, studies frequently use generic measures of health-related quality of life or health status in order to estimate quality-adjusted life years (QALYs) and the cost per QALY gained from one intervention compared to another. Alternative ways of estimating QALY weights exist, including the standard gamble and the time trade-off methods [2]. These can be complex to carry out within a trial and a common solution is to use a multi-attribute scale, which defines specific health states to which existing utility weights can be attached. The EuroQol/EQ-5D instrument is one such scale and has been used in numerous evaluations [3]. An alternative is the SF-6D, which is based on the SF-36 and is growing in popularity [4].

The sensitivity of utility scores derived from the EQ-5D and SF-6D to change in different disorders is unclear. In mental health there is concern over the use of such measures [5]. One of the likely limitations of existing QALY measures is their large focus on physical health problems. By concentrating on physical health it becomes much more of a challenge for mental health interventions to achieve a particular QALY gain compared with interventions for other conditions. This in turn could result in artificially high cost-utility ratios for mental health interventions. In the extreme case this could mean that resources are diverted away from mental health care by decision makers simply as a result of insensitive QALY measures. It is therefore crucial to ascertain whether these measures are sensitive to change in mental health and this paper seeks to compare the use of the EQ-5D and SF-6D using data from an evaluation of interventions for schizophrenia.

The specific aims of this study are to: (i) assess the relationship between SF-6D and EQ-5D utility scores for patients with schizophrenia at two points in time, (ii) assess the relationship in the change scores of these two measures, and (iii) measure the sensitivity of these measures to changes in an established measure of psychiatric symptoms.

Schizophrenia frequently follows a chronic course, and often results in high levels of disability [6], and this may well have an impact on quality of life. A small number of studies have examined the use of the EQ-5D in patients with schizophrenia, but few have assessed its validity [7]. We are not aware of any evaluations in schizophrenia that have used the SF-6D. Different methods of measuring utility can produce very different results [8] and it is important, therefore, to select the most suitable instrument. To inform this decision it is helpful to compare the sensitivity to change of different measures and this paper seeks to do this for the EQ-5D and SF-6D. One study has compared these two measures for patients with mood and anxiety disorders [9]. This found that both instruments were sensitive to differences in symptom severity but that the EQ-5D showed greater health gains. The study here uses similar methods but examines the use of the EQ-5D and SF-6D for a condition which is more severe and for which cost-utility analyses may be less easy to conduct.

Methods

Data for this study were collected as part of a multi-site European study which compared the impact that medication adherence therapy and a health education intervention for schizophrenia had on quality of life (the QUATRO study). Details of the study are provided elsewhere [10]. Patients with schizophrenia were recruited from the Netherlands, Germany,

the UK and Italy and were randomised to receive one of the two interventions. For inclusion, a clinical diagnosis of schizophrenia was confirmed using the Item Group Checklist of the Schedule for Clinical Assessment in Neuropsychiatry [11]. Recruitment was from general adult psychiatric settings (community and inpatient). A number of measures were taken at baseline and one-year follow-up and of particular relevance to the analyses presented here were the EQ-5D [3], the SF-36 [12] (which was the primary outcome measure in the clinical trial) and the 24-item Brief Psychiatric rating Scale (BPRS) [13]. Disability was measured with the Global Assessment of Functioning (GAF) [14].

The EQ-5D is a five dimensional questionnaire, the dimensions being mobility, self-care, usual activities, pain/discomfort and depression/anxiety. Each of these dimensions can be rated as 1 (no problem), 2 (some problem) or 3 (major problem). There are 243 distinct health states produced by the EQ-5D, and in the UK a sample of these have been valued by a community sample using the time trade-off method, with regression analyses used to estimate utility values for the remaining states [15].

Brazier and colleagues developed the SF-6D from the SF-36 [4]. The original instrument was reduced to six items (physical functioning, role limitations, social functioning, pain, mental health, vitality) each of which has between two and six levels of severity. This resulted in a total of 18,000 different health states and a sample of 249 of these were rated among 611 members of a community sample using the standard gamble method in order to generate preference-based utility scores. As with the EQ-5D, regression models were used to estimate utility scores for the remaining health states.

The BPRS is one of the most frequently used measures of psychiatric symptomatology. The expanded 24-item scale was used. This is an interviewer administered instrument and contains items relating to psychiatric symptoms during the past two weeks and also symptoms shown during the interview. Each item is rated between one (symptom not present) and seven (extremely severe symptom). A total score is obtained by summing scores for individual items, and a higher score indicates worse problems. This version of the BPRS has been demonstrated to have good inter-rater reliability [16] and to be sensitive to change [17]. Furthermore, Ruggeri et al have identified a common component structure for the instrument across five European sites (including three that are also included here) [18].

Data analytic procedures

EQ-5D and SF-6D utility scores (using UK weights) were computed and compared, with paired t-tests and the Wilcoxon test of differences in the mean ranks of observations used to identify significant differences (defined as $p<0.05$). Spearman's correlations were computed for both baseline and follow-up scores, and also for the change scores in both measures between the two time points.

Spearman's correlations that the EQ-5D and SF-6D had with the BPRS at baseline were computed. The sensitivity of the EQ-5D and SF-6D to changes in the BPRS was assessed by generating a Spearman's correlation of the change scores. In addition, we estimated standardised response mean (SRM) scores for the EQ-5D and SF-6D for patients who improved or deteriorated by at least 25% or on the BPRS. SRMs are similar to effect sizes and defined as the change between baseline and follow-up divided by the standard deviation of

this change [19]. Whilst any cut-off point is somewhat arbitrary, 25% is likely to represent minimal change on the BPRS that is clinically important [20].

Results

A total of 409 patients were randomised. There were slightly more men than women and three-quarters were White European (Table 1). Around one-third of the sample had been admitted to hospital during the previous year. The data on duration of time on neuroleptics and the GAF score indicate the chronicity and high disability of the sample. The mean (median) follow-up period was 373 (368) days with a range of 287-629. Complete EQ-5D ratings were available for 394 patients at baseline, 368 at follow-up and 358 at both time points. The respective figures for the SF-6D were 383, 367 and 345.

The mean EQ-5D and SF-6D utility scores were similar at both baseline and follow-up (Table 2). The difference of 0.01 in mean utility scores at baseline was not statistically significant (p=0.420). The difference at follow-up was 0.03 and this was significant (p=0.030). However, this difference is small and may not be clinically important. The median scores were substantially higher for the EQ-5D and the Wilcoxon test revealed that the distributions were significantly different at baseline (p=0.001), as can clearly be seen from Figures 1 and 2. Follow-up distributions (not shown) revealed similar differences (p<0.001). Variation as measured by the standard deviation was noticeably greater for the EQ-5D, and this is also shown by the range in scores. The median score for the SF-6D was similar to the mean, whilst for the EQ-5D it was higher. This indicates greater skewness in the EQ-5D scores than those

from the SF-6D. Patients scoring 1.000 on the EQ-5D comprised 17.3% of the sample at baseline and 20.9% at follow-up.

At baseline the Spearman correlation between the EQ-5D and SF-6D was 0.569 and 0.647 at follow-up. Both of these correlations were statistically significant (p<0.001). (These correlations being significant is to be expected given the relatively large sample size.)

The EQ-5D and SF-6D change scores showed a more modest relationship, with a Spearman correlation of 0.366 (p<0.001). The change scores for both the EQ-5D and SF-6D followed a symmetrical distribution, but there was a higher level of kurtosis for the EQ-5D (1.72) compared to the SF-6D (0.66) due to excess clustering at the mid point (charts not shown).

At baseline the correlation that the BPRS had with the EQ-5D and SF-6D was -0.343 and -0.314 respectively. There was a modest correlation between the change score on the EQ-5D and the BPRS (-0.29). A similar figure was observed for the SF-6D were similar (-0.22 BPRS). SRMs are shown in Table 3. The EQ-5D and SF-6D had identical scores for patients who improved by at least 25% on the BPRS. The figure of 0.39 would usually be interpreted as representing a moderate change. For patients who deteriorated by at least 25% the SRM was rather less for the EQ-5D. Negligible SRMs were seen for patients having a change of less than 25% on the BPRS.

Discussion

This is the first study we are aware of that has compared the EQ-5D and SF-6D in patients with schizophrenia. Utility scores derived using both approaches were highly correlated and mean scores were similar, even though the difference in mean scores at follow-up was statistically significant. However, the distributions were very different, with substantially higher median scores for the EQ-5D. The SF-6D utility scores followed a normal distribution, unlike the EQ-5D which were negatively skewed with a clustering at the maximum score of 1.000. This may mean that the two instruments are actually measuring different constructs or, at the very least, different aspects of quality of life. It was clear from the analysis that there was a ceiling effect for the EQ-5D, a finding that has been reported elsewhere [7]. The sample consisted of patients with schizophrenia, many of whom had been in contact with services for some years and had high levels of disability (shown by the GAF score in Table 1). This suggests that the EQ-5D, which indicated 17.3% having maximum quality of life at baseline and 20.9% at follow-up, does not reflect the severe nature of schizophrenia. This represents a clear limitation of this measure.

There was a significant but small relationship between EQ-5D and SF-6D change scores. The change in SF-6D utility scores followed a relatively normal distribution whilst change scores for the EQ-5D exhibited a high level of kurtosis. In terms of sensitivity to change in symptoms, both measures showed moderate change for patients who improved by at least 25% on the BPRS; however for those who had a worsening of symptoms, the SF-6D was slightly more sensitive. In another study, Stant et al found the EQ-5D to be insensitive to change in schizophrenia when compared to other outcome measures, but these did not include the SF-6D [1].

There are a number of limitations to this study. First, the analyses were performed on a pooled sample drawn from four European cities. This has the advantage of allowing the analyses to be performed on a large sample, but the relationships identified here may differ by site. Whilst we had the statistical power to test for site differences we did not do this as we did not wish to detract from the central message of the paper. Second, we used UK weights for the EQ-5D and SF-6D health states. Again, these may differ by site. Third, change in the BPRS total score was compared to change in the EQ-5D and SF-6D but certain subscales may have been more appropriate (for example changes in negative symptoms).

In conclusion we have seen that the EQ-5D and SF-6D produce similar mean utility scores in patients with schizophrenia. For the purposes of statistical analysis, on the basis of the distributions of actual utility scores and change scores the SF-6D has certain advantages. In particular, the EQ-5D appears to have a definite ceiling effect. On the basis of this study we would – cautiously – recommend the use of the SF-6D out of the two measures but further comparisons with the EQ-5D are also required.

References

1. Stant AD, Buskens E, Jenner JA, Wiersma D, TenVergert EM. Cost-effectiveness analysis in severe mental illness: outcome measures selection. *Journal of Mental Health Policy and Economics* 2007; **10:** 101-108.

2. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ* 1986; **5:** 1-30.

3. Williams A. *The Role of the EUROQOL Instrument in QALY Calculations.* York: Centre for Health Economics, University of York, 1995.

4. Brazier JE, Roberts JF, Deverill MD. The estimation of a preference based measure of health from the SF-36**.** *J Health Econ* 2002; **21:** 271-292.

5. Chisholm D, Healey A, Knapp M. QALYs and mental health care. *Soc Psychiatry Psychiatr Epidemiol* 1997; **32:** 68-75.

6. Harvard School of Public Health, World Health Organisation, World Bank. *Global Burden of Disease*, 1996.

7. König H-H, Roick C, Angermeyer MC. Validity of the EQ-5D in assessing and valuing health status in patients with schizophrenic, schizotypal or delusional disorders. *Eur Psychiatry* 2007; **22:** 177-187.

8. Marra CA, Marion SA, Guh DP, Najafzadeh M, Wolfe F, Esdaile JM, Clarke AE, Gignac MA, Anis AH. Not all "quality-adjusted life years" are equal. *Journal of Clinical Epidemiology* 2007; **60:** 616-624.

9. Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Economics* 2006; **15:** 1229-1236.

10. Gray R, Leese M, Bindman J, Becker T, Burti L, David A, Gournay K, Kikkert M, Koeter M, Puschner B, Schene A, Thornicroft G, Tansella M. Adherence therapy for people with schizophrenia: European multicentre randomised controlled trial. *Br J Psychiatry* 2006; **189:** 508-514.

11. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 1990; **47:** 589-593.

12. Ware J, Sherbourn C. The MOS, 36-item Short-Form Health Survey (SF-36). I Conceptual framework and item selection. *Med Care* 1992; **30:** 473-483.

13. Lukoff D, Neuchterlein KH, Ventura J. Manual for the expanded Brief Psychiatric Rating Scale (BPRS). *Schizophrenia Bulletin* 1986; **12:** 594-602.

14. Endicott J, Spitzer R, Fleiss J, Cohen J. The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry* 1976; **33:** 766-771.

15. Dolan P, Gudex C, Kind P, Williams A (1995). *A Social Tariff for EuroQol: Results from a UK General Population Survey.* York: Centre for Health Economics, University of York.

16. Ventura J, Green M, Shaner A, Liberman R. Training and quality assurance with Brief Psychiatric Rating Scale: "The Drift Busters." *International Journal of Methods in Psychiatric Research* 1993; **3:** 221-244.

17. Burlingame GM, Seaman S, Johnson JE, Whipple J, Richardson E, Rees F, Earnshaw D, Spencer R, Payne M, O'Neil B. Sensitivity to change of the Brief Psychiatric Rating Scale-Extended (BPRS-E): an item and subscale analysis. *Psychological Services* 2006; **3:** 77-87.

18. Ruggeri M, Koeter M, Schene A, Bonetto C, Vazquez-Barquero JL, Becker T, Knapp M, Knudsen HC, Tansella M, Thornicroft G. Factor solution of the BPRS-expanded version in schizophrenic outpatients living in five European countries. *Schizophrenia Research* 2005; **75:** 107-117.

19. Löwe B, Kroenke K, Herzog W Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord* 2004; **81:** 61-66.

20. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief Psychiatric Rating Scale scores. *Br J Psychiatry* 2005; **187:** 366-371.

Table 1. Patient characteristics at baseline.

Characteristics	Valid number of cases	Mean (sd)
Age	409	41.5 (11.5)
Years on medication	366	13.7 (9.9)
Psychiatric in-patient days in last year	409	27.9 (60.4)
GAF total score	407	50.3 (13.9)
		N (%)
Male	409	245 (59.9%)
White European	409	310 (75.8%)
Any psychiatric inpatient admission in last year	409	159 (38.9%)

GAF = Global Assessment of Functioning

Table 2. EQ-5D, SF-6D, MANSA and BPRS scores at baseline and follow-up.

	N	Mean	SD	Median	Min	Max
EQ-5D baseline	394	0.68	0.29	0.75	-0.35	1.00
EQ-5D follow-up	368	0.71	0.29	0.80	-0.48	1.00
EQ-5D change	358	0.04	0.29	0.00	-1.10	0.87
SF-6D baseline	383	0.67	0.13	0.66	0.32	1.00
SF-6D follow-up	367	0.68	0.13	0.68	0.30	1.00
SF-6D change	345	0.01	0.12	0.01	-0.32	0.41
MANSA baseline	408	4.45	0.98	4.50	1.25	6.75
MANSA follow-up	370	4.58	0.99	4.63	1.00	7.00
MANSA change	369	0.12	0.91	0.06	-2.88	3.31
BPRS baseline	406	45.17	13.02	43.00	24.00	103.00
BPRS follow-up	371	37.71	10.54	35.00	24.00	84.00
BPRS change	369	-7.38	12.82	-7.00	-62.00	28.00

Table 3. Standardised response mean (SRM)[a] scores of EQ-5D and SF-6D by level of change on BPRS.

BPRS score	EQ-5D	SF-6D
Improvement of at least 25%	0.39 (n=119)	0.39 (n=116)
Deterioration/improvement of less than 25%	0.05 (n=202)	0.02 (n=194)
Deterioration of at least 25%	-0.17 (n=30)	-0.27 (n=29)

[a] Standardised response mean is defined as the change between baseline and follow-up divided by the standard deviation of this change.

Figure 1. Distribution of EQ-5D utility scores at baseline.
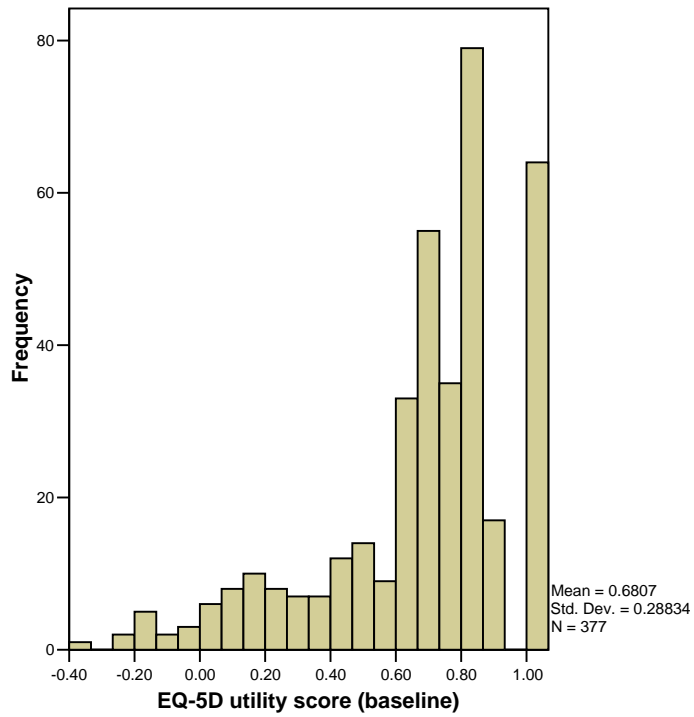

Mean = 0.6807
Std. Dev. = 0.28834
N = 377

Figure 2. Distribution of SF-6D utility scores at baseline.

Mean = 0.6707
Std. Dev. = 0.12323
N = 377