

Measuring the benefits of treatment for psychosis: validity and responsiveness of the EQ-5D

Garry R. Barton, Jo Hodgekins, Miranda Mugford, Peter B. Jones, Tim Croudace and David Fowler

Background

The UK National Institute for Health and Clinical Excellence (NICE) has recommended that cost-effectiveness analysis includes the EQ-5D; however, this is often not implemented in the area of mental health.

Aims

To assess the appropriateness of using the EQ-5D to measure improvements in mental health.

Method

Seventy-seven participants with psychosis were rated according to the EQ-5D and seven measures of mental health at both pre- and post-intervention. To assess construct validity we compared the (pre-intervention) mean EQ-5D scores for those with milder and more severe scores, according to each of the seven measures. To assess

responsiveness we estimated the mean EQ-5D change score for those who improved (post-intervention), according to each of the measures.

Results

The mean EQ-5D score was more favourable for both those with milder scores (mean difference: 0.044 to 0.301) and for those who improved post-intervention (mean change: 0.029 to 0.117).

Conclusions

This suggests the EQ-5D should be considered for use in future cost-effectiveness studies in the area of mental health.

Declaration of interest

None.

Cost-effectiveness studies conducted alongside interventions aimed at improving mental health are often limited in scope because the original study omitted to include a utility measure.^{1,2} A utility measure provides a common scale on which to compare the benefits of different interventions,^{3,4} where zero is equivalent to death and one is equivalent to full health. Without such a measure, it is often difficult to conclude whether the intervention in question represents a cost-effective use of scarce resources, or whether resources would be better spent elsewhere. This is highlighted by a number of recent cost-effectiveness studies in the area of mental health^{5–11} that have found an intervention to be both more costly and more effective, but as effectiveness was not measured in terms of utility it was not possible to compare the cost-effectiveness of these interventions with that for other healthcare interventions. Given the inferred reluctance to use the EQ-5D when evaluating mental health interventions, and the possibility that the EQ-5D may not be sensitive to changes in quality of life in this area,¹² in this paper we seek to assess the appropriateness of using the EQ-5D to measure the benefits of providing social recovery oriented cognitive-behavioural therapy.

Akin to analyses undertaken in other clinical areas, for example Hurst *et al*¹³ and Terwee *et al*,¹⁴ we thereby assessed the validity and responsiveness of the EQ-5D in a group of people with psychosis. The importance of such an assessment is highlighted by the fact that the UK National Institute for Health and Clinical Excellence (NICE) has recently stated that the EQ-5D is the preferred measure to be used in cost-effectiveness analyses.¹⁵

ISREP trial was designed to compare the effectiveness and cost-effectiveness of two interventions – case management alone and social recovery oriented cognitive-behavioural therapy in addition to case management, where social recovery oriented cognitive-behavioural therapy was available for a 9-month period post-randomisation. Ethical approval was granted by the Norfolk local research ethics committee and participants in this study were recruited from two secondary mental health services. The inclusion criteria was: a current diagnosis of affective or non-affective psychosis (including schizophrenia, schizoaffective disorder, bipolar disorder and psychotic depression); illness duration ≤ 8 years; positive psychotic symptoms (hallucinations and delusions) in relative remission (denoted by a score of ≤ 4 on individual symptoms on the Positive and Negative Syndrome Scale (PANSS));¹⁷ and currently unemployed or engaged in less than 16 h paid employment or education. Participants were excluded if the psychotic disorder was thought to have an organic basis, acute psychosis was present or the primary diagnosis was drug dependency on opiates or cocaine.

Outcome measures

Participants in the ISREP trial were rated according to seven measures of mental health, and the EQ-5D, at both baseline and 9 months post-randomisation (9-month assessment). Five of these seven measures (the Beck Anxiety Inventory (BAI),¹⁸ Beck Depression Inventory (BDI),¹⁹ Beck Hopelessness Scale (BHS),²⁰ PANSS¹⁷ and Global Assessment of Functioning Scale (GAF; symptom ratings only))²¹ aim to capture the severity of various mental health symptoms, whereas the remaining two (the Quality of Life Scale (QLS),²² and Social and Occupational Functioning Assessment Scale (SOFAS)²³) focus more on the level of functioning. The BAI, BDI, BHS and EQ-5D were completed by the study participants, whereas the PANSS, QLS, SOFAS and GAF were rated by a member of the ISREP study team.

Method

Participants

All participants were taking part in the Improving Social Recovery in Early Psychosis (ISREP) trial. The methods of this study have been outlined elsewhere;¹⁶ briefly they were as follows. The

The BAI assesses the extent to which an individual is bothered by a particular symptom¹⁸ and the BDI assesses the intensity of a particular depressive symptom.¹⁹ Both the BAI and BDI consist of 21 items which are scored on a 0–3 scale, where a higher score denotes more severe symptoms. The BHS consists of 20 items and is designed to assess the degree to which an individual holds negative perceptions about the future.²⁰ A true or false response format is used and a higher score denotes more negative perceptions. The PANSS assesses the levels of positive, negative and cognitive symptoms that are associated with psychosis,¹⁷ where 30 items are scored on a 1 (absent) to 7 (extreme) scale and a higher score reflects a greater psychopathology. The QLS is designed to assess the functional impairments associated with psychosis,²² where 21 items are assessed on a 0–6 scale and high scores reflect normal or unimpaired function. The GAF assesses symptom level and psychological, social and occupational functioning,²¹ and the SOFAS measures social and occupational functioning.²³ Both measures are assessed on a 1–100 rating scale that is divided into ten deciles, each of which provides a description of functioning and symptom level. A lower score on both the GAF and SOFAS denotes a worse response.

The EQ-5D was developed by the EuroQol group (a consortium of researchers from Western Europe) in the 1990s and was informed by a survey of lay people's concepts of health, and a review of existing instruments.²⁴ There are five questions, where the respondent is asked to report the level of problems they have (no problems, some/moderate problems, and severe/extreme problems) with regard to mobility, self-care, usual activities, pain, and anxiety/depression.²⁵ Responses to these five dimensions are also converted into one of 243 different EQ-5D health state descriptions, which range between no problems on all five dimensions (11111) and severe/extreme problems on all five dimensions (33333). We assigned a utility score to each of these 243 health states using the York A1 tariff,²⁶ giving a measure of health status which ranges between –0.594 and 1 (full health), and where death is equal to zero. This tariff was based on the preferences elicited from a survey of 3395 members of the UK population who used the time trade-off technique to value a number of potential EQ-5D states (the time trade-off seeks to establish how much one would be willing to reduce one's life expectancy by in order to obtain full health).²⁴ It should thereby be noted that very few of those who valued the EQ-5D states will actually have experienced severe mental illness. The potential importance of this is highlighted by the fact that members of the public tend to estimate the loss in utility associated with particular health states to be higher than that reported by individuals who have actually experienced the health states in question.²⁷ However, the argument for using valuations based on the general public is that, in a publicly funded system, the views of the general public are most relevant.²⁸

Analyses

Overview

Although NICE stated that the EQ-5D was its preferred measure to be used in cost-effectiveness analyses, it did note that the EQ-5D may not be appropriate in all circumstances.¹⁵ Providing evidence of the latter is however not straight forward given that there is much conceptual and methodological difficulty associated with the decision as to whether a measure is 'appropriate',²⁹ and myriad of differently defined terms have been used within such assessments.³⁰ That said, Fitzpatrick *et al*³¹ outlined a number of criteria on which evidence should be provided in order to select an appropriate outcome measure. Similarly, Brazier *et al*³² developed a checklist for judging the merits of preference-based

measures of health such as the EQ-5D. Informed by these papers, and the fact that very few papers have assessed utility measures with regard to such criteria,³³ we measured the performance of the EQ-5D with regard to the criteria of construct validity, convergent validity and responsiveness.

Validity

Validity was assessed in terms of both construct and convergent validity. Construct validity relates to whether a measure can discriminate between two patient groups, one which has a certain trait, and the other which does not.³⁰ Thus, we sought to assess whether participants in the ISREP trial, who had what might be considered milder scores (at baseline) according to each of the seven mental health measures, had different EQ-5D scores to those who might be considered to have more severe scores. For each of the seven mental health measures, participants were thereby categorised into two groups using the following methods.

On the BAI it has been recommended that scores of 0–9 points be interpreted as normal anxiety, 10–18 as mild–moderate, 19–29 as moderate–severe, and 30–63 as severe anxiety.¹⁸ Thus, we developed two groups – those who had BAI scores ≤ 18 and those who had BAI scores ≥ 19 . On the BDI it has been recommended that scores of 0–13 correspond to minimal depression, 14–19 as mild depression, 20–28 as moderate depression, and 29–63 as severe depression.³⁴ Accordingly, two groups of participants were created – BDI scores ≤ 19 and BDI scores ≥ 20 . On the BHS a score of 0–3 can be considered minimal, 4–8 as mild, 9–14 as moderate, and 15–20 as severe.²⁰ Therefore, we compared those with a BHS score ≤ 8 with those who had a score ≥ 9 . On the GAF, a score in the range 51–60 is said to denote moderate symptoms or moderate difficulty in functioning, whereas a score of 61–70 denotes some mild symptoms or some difficulty in functioning.²³ Similarly, on the SOFAS, a score of 51–60 is said to denote moderate difficulty in functioning, whereas a score of 61–70 denotes some difficulty in functioning.²³ Accordingly, for both the GAF and SOFAS, those with scores ≤ 60 were compared with those with scores ≥ 61 . For the remaining two measures we were unable to identify a range on the respective total scores that could be considered to denote mild symptoms (on the PANSS) or a mild level of functional impairment (on the QLS). Consequently, for these two measures, in an attempt to compare those who had milder scores with those who had more severe scores, we simply split the sample into approximately two equal sized groups and compared those who had higher scores with those who had lower scores. Finally, where the aforementioned splits (for either the BAI, BDI, BHS, SOFAS or GAF) resulted in particularly unequal numbers in the milder/more severe groups, the same analysis was also performed using a different split that resulted in approximately equal numbers in each of the two groups.

In order to compare the mean EQ-5D score for participants in each of the above mild/more severe groups, using baseline scores, we conducted the *t*-test to assess whether the EQ-5D could discriminate between the two groups of participants (who had different scores according to each of the seven measures of mental health). However, as the *t*-test requires the data to be approximately normally distributed (this was assessed using the Kolmogorov–Smirnov *Z*-test) and it has been shown that responses to the EQ-5D do not tend to be normally distributed,³⁵ we also compared the EQ-5D scores for both the mild/more severe groups using the Mann–Whitney *U*-test. Both the *t*-test and *U*-test were conducted in order to assess the hypothesis that non-parametric and parametric methods produce similar results and that the latter are thereby robust to the violation of the

normality assumption that health-related quality-of-life data are likely to cause.³⁶ Finally, it should be noted that more complex regression approaches were considered superfluous as, in contrast to other studies, for example Hurst *et al*¹³ who recruited consecutive out-patients, the strict entry criteria to this study reduced the need to adjust for other factors which might differ between the two groups of participants.

We also assessed whether the mean difference between the EQ-5D scores, at baseline, in different groups could be considered to constitute a minimally important difference (MID). The MID is considered to be the smallest change in score that would necessitate a change in a person's management.³⁷ Previously the MID has, for example, been estimated by calculating the mean change in the EQ-5D score for those who reported that their general health was either somewhat better than a year ago, or somewhat worse than a year ago, in response to question two of the Short Form-36 (SF-36).³⁷ Here, in line with the assumption made in a previous paper³⁸ that considered estimates of the MID which were reported in a number of studies, we assumed that a difference of >0.03 constituted a MID on the EQ-5D.

Finally, as previously undertaken by Kind *et al*,³⁹ in an attempt to explain the different EQ-5D scores we also estimated the proportion of participants who reported that they had a problem on each of the five dimensions of the EQ-5D. Here the chi-squared test was also conducted in order to assess whether the proportion of participants reporting a particular problem also differed between those with milder scores and those with more severe scores. Convergent validity is determined by how closely a measure is related to other measures of the same construct.³⁰ Thus, we used the Spearman rank test to assess whether (baseline) scores on the EQ-5D were correlated with the scores on each of the seven measures of mental health in the direction that one would expect.

Responsiveness

Responsiveness has been defined as the ability of a scale to detect changes.⁴⁰ In order to assess this criterion, we estimated the mean change in the EQ-5D score (between baseline and follow-up) for those whose scores improved post-intervention, according to each of the seven mental health measures, and compared it with the mean change for those whose scores did not improve (i.e. those whose score worsened or remained the same post-intervention). This definition of responsiveness was chosen in preference to others (see e.g. Terwee *et al*¹⁴) as it is in line with the argument made by Claxton⁴¹ – that decisions (with regard to cost-effectiveness) should be made on the basis of mean values, irrespective of whether such differences are considered for example clinically important or statistically significant. Nevertheless, for the reasons

outlined previously, to assess the significance of any differences between these two groups we again conducted the *t*-test and the *U*-test. Additionally, the mean difference between these two groups was assessed in relation to the assumed EQ-5D MID of 0.03. Finally, the change in the proportion of participants who reported having no problems (post-intervention compared with pre-intervention on each of the five dimensions of the EQ-5D) was calculated for both those who improved and did not improve post-intervention according to each of the seven measures of mental health.

Results

Participants

Within the two secondary health centres in question between January 2005 and July 2006 a total of 200 participants were identified as meeting the aforementioned inclusion criteria. Of these, 88 gave consent to take part in the study, although a further 11 of these dropped out during the baseline assessment (5 became symptomatic, 5 stated that they were no longer interested, and 1 for personal reasons). Thus, 77 participants were recruited into the study, 55 (71.4%) of whom were male, 50 (64.9%) of whom had a diagnosis of non-affective psychosis, and the average age was 28.9 years (range 18–52). The average duration of illness and the average length of unemployment were 4.8 years and 209 weeks respectively.

At baseline, the EQ-5D was completed by 68 participants (88.3%), where the mean score was 0.676 (95% CI 0.604–0.748) compared with 0.743 (95% CI 0.671–0.816) at 9 months, giving a mean change of 0.043 (95% CI –0.034 to 0.122) (see Table 1 where the mean scores for the BAI, BDI, BHS, PANSS, QLS, GAF and SOFAS are also summarised). Across the five dimensions the proportion who reported a problem on each of the dimensions of the EQ-5D was 26.5% (mobility), 22.1% (self-care), 51.5% (usual activities), 39.7% (pain/discomfort) and 70.6% (anxiety/depression) at baseline ($n=68$), compared with 18.8%, 12.5%, 43.8%, 31.3% and 66.7% respectively, at 9 months post-intervention ($n=48$).

Analyses

Construct validity

The two groups of participants that were created according to scores on each of the seven measures of mental health had approximately equal numbers, with the exception of the SOFAS. Thus, as well as comparing the EQ-5D scores for those with a SOFAS score of ≥ 61 ($n=7$) with those with a score ≤ 60 ($n=61$), we also compared the EQ-5D scores for those with a SOFAS score ≥ 51 ($n=25$) with those with a score ≤ 50

Table 1 Mean scores at both baseline and 9 months post-intervention, and the change in score for each of the seven measures of mental health and the EQ-5D

	Baseline score, mean (<i>n</i>)	9-month score, mean (<i>n</i>)	Change, mean (<i>n</i>)
Beck Anxiety Inventory	16.97 (74)	13.11 (62)	–3.50 (62)
Beck Depression Inventory	21.90 (73)	14.05 (59)	–8.15 (59)
Beck Hopelessness Scale	8.80 (74)	7.26 (57)	–1.21 (56)
Positive and Negative Syndrome Scale	56.74 (77)	50.42 (62)	–6.30 (62)
Quality of Life Scale	64.54 (76)	74.13 (63)	8.96 (63)
Global Assessment of Functioning Scale ^a	56.83 (77)	59.77 (70)	2.73 (70)
Social and Occupational Functioning Assessment Scale	50.06 (77)	54.25 (69)	3.78 (69)
EQ-5D	0.676 (68)	0.743 (48)	0.043 (45)

a. Symptom ratings only.

Table 2 Baseline mean EQ-5D scores for those with milder and more severe scores and the mean difference between these two groups

	EQ-5D score, mean (n)	Difference, mean (95% CI)	
Beck Anxiety Inventory ≤ 18 ≥ 19 (more severe)	0.788 (41) 0.487 (26)	0.301 ^{***‡} (0.171 to 0.431)	
Beck Depression Inventory ≤ 19 ≥ 20 (more severe)	0.821 (30) 0.540 (36)	0.282 ^{***‡} (0.151 to 0.412)	
Beck Hopelessness Scale ≤ 8 ≥ 9 (more severe)	0.794 (36) 0.538 (31)	0.255 ^{***††} (0.122 to 0.389)	
Positive and Negative Syndrome Scale ≤ 55 ≥ 56 (more severe)	0.712 (35) 0.638 (33)	0.074 (−0.070 to 0.219)	
Quality of Life Scale ≥ 66 ≤ 65 (more severe)	0.697 (31) 0.652 (35)	0.044 (−0.105 to 0.193)	
Global Assessment of Functioning Scale ^a ≥ 61 ≤ 60 (more severe)	0.759 (24) 0.631 (44)	0.129 (−0.020 to 0.277)	
Social and Occupational Functioning Assessment Scale ≥ 61 ≤ 60 (more severe)	0.787 (7) 0.663 (61)	0.124 (−0.113 to 0.361)	
	≥ 51 ≤ 50 (more severe)	0.702 (25) 0.661 (43)	0.042 (−0.109 to 0.192)

a. Symptom ratings only.
* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ according to the t -test and † $P < 0.05$, †† $P < 0.01$, ‡ $P < 0.001$ according to the U -test.

($n = 43$) (a score of 41–50 on the SOFAS denotes serious impairment in social, occupational or school functioning).²³ For each of the seven measures of mental health (including both SOFAS categorisations), those with milder scores were found to have higher mean scores according to the EQ-5D compared with those with more severe scores (Table 2). However, according to the

Kolmogorov–Smirnov Z -test ($z = 1.964$, $P < 0.001$, $n = 68$), one would reject the null hypothesis that the EQ-5D data were normally distributed. Thus, the significance levels according to the t -test, with regard to the mean difference in utility between those with milder/more severe scores, should be treated with caution, although they are near identical to the significance levels

Table 3 Baseline proportion who reported having a problem on each of the five dimensions of the EQ-5D for those with milder and more severe scores

	n	Mobility, %	Self-care, %	Usual activities, %	Pain/discomfort, %	Anxiety/depression, %
Beck Anxiety Inventory ≤ 18 ≥ 19 (more severe)	41 26	17.07 42.31*	17.07 30.77	39.02 73.08**	29.27 57.69*	58.54 92.31**
Beck Depression Inventory ≤ 19 ≥ 20 (more severe)	30 36	23.33 30.56	13.33 30.56	30.00 69.44**	23.33 55.56**	53.33 88.89**
Beck Hopelessness Scale ≤ 8 ≥ 9 (more severe)	36 31	16.67 38.71*	8.71 38.71**	36.11 67.74**	30.55 48.39	58.33 83.87*
Positive and Negative Syndrome Scale ≤ 55 ≥ 56 (more severe)	35 33	22.86 30.30	22.86 21.21	45.71 57.58	40.00 39.39	68.57 72.73
Quality of Life Scale ≥ 66 ≤ 65 (more severe)	35 31	22.58 28.57	22.58 22.86	51.61 51.43	35.48 42.86	67.74 74.29
Global Assessment of Functioning Scale ^a ≤ 60 ≥ 61 (more severe)	24 44	20.83 29.55	16.67 25.00	50.00 52.27	25.00 47.73	70.83 70.45
Social and Occupational Functioning Assessment Scale ≤ 60 ≥ 61 (more severe)	7 61	28.57 26.22	14.29 22.95	42.86 52.46	42.86 39.34	57.14 72.13
	25 43	24.00 27.91	20.00 23.26	56.00 48.84	40.00 39.54	28.00 82.09

a. Symptom ratings only.
* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

according to the *U*-test (Table 2). Additionally, it should be noted that the mean difference exceeded the assumed MID of 0.03 for all seven measures (range 0.044–0.301).

Similar results were also achieved on each of the five dimensions of the EQ–5D. In Table 3 it can be seen that the proportion of participants who reported having problems with each of the dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression was lower for those with milder scores compared with those with more severe scores according to at least five of the seven measures of mental health.

In terms of convergent validity the EQ–5D was correlated with each of the seven measures of mental health in the direction that one would expect, i.e. it was negatively correlated with those measures for which a lower score denotes more severe symptoms/functioning (the BAI, BDI, BHS and PANSS) and positively correlated with the QLS, GAF and SOFAS (Table 4). However, it should be noted that the level of correlation was not significant for three of the measures (the PANSS, QLS and SOFAS).

Table 4 Baseline levels of correlations between each of the seven measures of mental health and the EQ–5D according to the Spearman rank test

	EQ–5D (n)
Beck Anxiety Inventory	–0.656 (67)***
Beck Depression Inventory	–0.360 (55)**
Beck Hopelessness Scale	–0.459 (67)***
Positive and Negative Syndrome Scale	–0.228 (68)
Quality of Life Scale	0.025 (68)
Global Assessment of Functioning Scale ^a	0.263 (68)*
Social and Occupational Functioning Assessment Scale	0.053 (68)

a. Symptom ratings only.
P* < 0.05, *P* < 0.01, ****P* < 0.001.

Responsiveness

In Table 5 it can be seen that the mean EQ–5D score was higher (post-intervention) for each of the participant groups who improved according to the seven measures of mental health, i.e. when scores on the BAI, BDI, BHS and QLS increased and scores on the PANSS, GAF and SOFAS decreased. However, contrary to expectations, those who did not improve according to the BHS actually achieved a larger increase in the mean EQ–5D score than was the case for the participants who actually had better BHS scores post-intervention (see Discussion for possible explanations). The significance levels associated with these mean differences are also denoted in Table 5 where, according to the Kolmogorov–Smirnov *Z*-test (*z* = 0.857, *P* = 0.455, *n* = 45), one cannot reject the null hypothesis that the EQ–5D change scores were normally distributed.

As for the individual dimensions of the EQ–5D (Table 6), it can be seen that of those who improved (post-intervention) according to each of the seven measures of mental health, a greater proportion tended to report having no problems (post-intervention compared with pre-intervention) with regard to each of the dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The main exceptions were that at the 9-month assessment (compared with the baseline) more of those who improved according to the PANSS (2.94%: *n* = 1 more) reported having a problem with regard to usual activities and more of those who improved according to the SOFAS (6.9%: *n* = 2 more) reported having a problem with regard to anxiety/depression.

Discussion

We have shown that the mean EQ–5D score for those who were considered to have milder scores was higher than that for those who had more severe scores, according to each of the BAI, BDI, BHS, PANSS, QLS, GAF and SOFAS, and that the mean difference was greater than the assumed MID of 0.03 (Table 2). The EQ–5D

Table 5 Mean change in EQ–5D scores for those whose scores improved, or did not improve, post-intervention and the mean difference between these two groups

	Change in EQ–5D score, mean (n)	Difference, mean (95% CI)
Change in Beck Anxiety Inventory		
≥ 0	–0.069 (17)	0.181*,† (0.027 to 0.335)
≤ –1 (improved score)	0.112 (28)	
Change in Beck Depression Inventory		
≥ 0	–0.089 (13)	0.179*,† (0.027 to 0.331)
≤ –1 (improved score)	0.090 (30)	
Change in Beck Hopelessness Scale		
≥ 0	0.032 (14)	–0.003 (–0.172 to 0.167)
≤ –1 (improved score)	0.029 (28)	
Change in Positive and Negative Syndrome Scale		
≥ 0	–0.029 (9)	0.103 (–0.096 to 0.301)
≤ –1 (improved score)	0.073 (34)	
Change in Quality of Life Scale		
≤ 0	0.012 (11)	0.044 (–0.142 to 0.232)
≥ 1 (improved score)	0.057 (31)	
Change in Global Assessment of Functioning Scale ^a		
≤ 0	–0.052 (19)	0.169*,† (0.014 to 0.324)
≥ 1 (improved score)	0.117 (25)	
Change in Social and Occupational Functioning Assessment Scale		
≤ 0	0.022 (15)	0.033 (–0.138 to 0.205)
≥ 1 (improved score)	0.056 (29)	

a. Symptom ratings only.
P* < 0.05, *P* < 0.01, ****P* < 0.001 according to the *t*-test and †*P* < 0.05, ‡*P* < 0.01, §*P* < 0.001 according to the *U*-test.

was also correlated with each of these seven measures of mental health, in that those with better scores on each of these dimensions also tended to have better EQ-5D scores, although the level of correlation was not significant for three of the measures (Table 4). Additionally, those who improved (post-intervention) according to each of the measures of mental health also had higher mean EQ-5D scores at the 9-month assessment compared with the baseline assessment (Table 5).

It can also be seen that those who had milder scores according to each of the seven measures of mental health, also tended to have fewer problems than those with more severe scores across each of the five dimensions of the EQ-5D (Table 3). Thus, this would suggest that many of the dimensions of the EQ-5D can be sensitive to changes in the level of mental health. Further support for this argument is provided by the fact that a greater proportion of those who improved, according to each of the seven measures of mental health, also tended to report having no problems on each of the five dimensions of the EQ-5D at follow-up compared with baseline (Table 6).

One seemingly unexpected result was, however, that those who did not improve according to the BHS actually had a higher mean post-intervention EQ-5D score than those who did improve. Looking at the results in Table 6, it can be seen that (post-intervention), as one might expect, a greater proportion of those who had more negative perceptions about the future also reported having problems with regard to the anxiety/depression dimension of the EQ-5D. However, a greater proportion also reported having no problems and 15 individuals whose perceptions about the future did not improve actually received other benefits (reflected in the mobility, self-care and pain/discomfort dimensions of the EQ-5D) that outweighed the increased levels of anxiety and depression, although we cannot account for why this occurred. In a similar way, the finding that a greater proportion of those who improved according to the SOFAS actually reported having problems with regard to anxiety and depression may not be contrary to expectations as the SOFAS focuses on social and occupational functioning. That said, the finding that the proportion of those who reported having a problem with regard

to usual activities increased among those who improved according to the PANSS cannot be fully accounted for. Similarly, the argument that our results are far from conclusive is bolstered by the fact that only four of the seven measures of mental health were significantly correlated with the EQ-5D.

Comparisons with other studies

A number of previous cost-effectiveness studies in the area of mental health have included the EQ-5D as a measure of benefit, for example Byford *et al.*,¹² Palmer *et al.*,⁴² Byford *et al.*⁴³ and Hakkaart-van Roijen *et al.*⁴⁴ However, in each of these studies,^{12,42–44} according to the EQ-5D, the intervention was estimated to be no more effective than the comparator with which it was compared (this was often termed ‘treatment as usual’). Such results could arise because there was an improvement associated with the intervention, but the EQ-5D may have been insensitive to that improvement, as was pointed out by Byford *et al.*¹² Alternatively, it could be that any benefits do not constitute sufficient value, according to the EQ-5D, in order for them to increase a person’s level of utility. Within this study we have tried to inform this debate by considering the validity and responsiveness of the EQ-5D in a group of people with psychosis. Our findings suggest that the EQ-5D can discriminate between those with milder and more severe scores, and that the EQ-5D is generally responsive to improvements in mental health as measured by the BAI, BDI, BHS, PANSS, QLS, GAF and SOFAS. These findings are consistent with others who have found evidence to support the validity and/or responsiveness of the EQ-5D in people with schizophrenic, schizotypal or delusional disorders,⁴⁵ in a group of individuals receiving mental health services⁴⁶ and in other areas of health.^{13,47–49} However, this does not necessarily imply that similar results will be realised in other groups with severe mental illness.

Readers should also be aware that, in addition to the EQ-5D, other measures of utility measurement are available, including the Health Utilities Index⁵⁰ and the SF-6D⁵¹ (the latter of which is derived from the SF-36).⁵² A number of studies have thereby been undertaken to compare the utility scores derived from these

Table 6 Change in the proportion who reported having no problems on each of the five dimensions of the EQ-5D (9-month compared with baseline assessment) for those whose scores improved, and did not improve, post-intervention

	<i>n</i>	Mobility %	Self-care %	Usual activities, %	Pain/discomfort %	Anxiety/ depression, %
Change in Beck Anxiety Inventory						
≥ 0	17	5.88	11.76	-17.65	5.88	-29.41
≤ -1 (improved score)	28	14.29	14.29	14.29	7.14	14.29
Change in Beck Depression Inventory						
≥ 0	13	7.69	-7.69	15.38	-7.69	23.08
≤ -1 (improved score)	30	16.67	13.33	13.33	6.67	10.00
Change in Beck Hopelessness Scale						
≥ 0	14	21.43	7.14	-21.43	14.29	-6.67
≤ -1 (improved score)	28	3.57	17.86	7.10	0.00	3.57
Change in Positive and Negative Syndrome Scale						
≥ 0	9	0.00	22.22	22.22	0.00	0.00
≤ -1 (improved score)	34	14.71	11.76	-2.94	11.76	0.00
Change in Quality of Life Scale						
$= 0$	11	9.09	18.18	-9.09	0.00	0.00
≥ 1 (improved score)	31	9.68	9.68	3.23	9.68	0.00
Change in Global Assessment of Functioning Scale ^a						
≤ 0	19	0.00	5.26	-10.53	-10.53	-10.53
≥ 1 (improved score)	25	20.00	20.00	12.00	20.00	4.00
Change in Social and Occupational Functioning Assessment Scale						
≤ 0	15	6.67	13.33	0.00	6.67	6.67
≥ 1 (improved score)	29	13.79	13.79	3.45	6.90	-6.90

a. Symptom ratings only.

different multi-attribute health status classification systems, where a common finding is that there are small but important differences between the utility scores estimated by each of the measures.⁵³ This is in accordance with the conclusion of the only paper which we know to have compared the scores provided by two utility measures (the EQ-5D and SF-6D) in the area of mental health.⁵⁴ In a group of people with either a major depressive disorder, dysthymic disorder, panic disorder, social phobia, or generalised anxiety, Lamers *et al* found that those who were less distressed (according to a checklist of 90 psychological symptoms) tended to have higher scores on the EQ-5D (compared with the SF-6D), whereas those who were more distressed tended to have higher scores on the SF-6D.⁵⁴ Differing scores such as these are often explained by the fact that different measures use both different health-state descriptions and different valuation techniques.²⁹

Limitations

The main limitation of this study is that it is relatively small (based on 77 participants), and no single experiment can unequivocally prove a construct (evidence of validity can only be provided by a series of converging results).³⁰ Additionally, although the *t*-test has previously been used to analyse responses to the EQ-5D (e.g. Hurst *et al*¹³) there may be some limitations with regard to the statistical analysis due to the fact that the EQ-5D data were not normally distributed (a requirement for the *t*-test). That said, we demonstrated that the results of the *t*-test were robust as the qualitative interpretation of these results (in relation to the $P < 0.05$ cut-off) was identical to those obtained using the Mann-Whitney *U*-test. Furthermore, it should also be recognised that we did not collect sufficient data to enable all of the criteria developed by both Fitzpatrick *et al*³¹ and Brazier *et al*³² to be fully assessed. Further research is therefore required before one can conclude that it is wholly appropriate to use the EQ-5D to measure the benefits in mental health evaluations. That said, our results suggest that, for this particular intervention (which focused on social recovery),¹⁶ those who improved according to the BAI, BDI, BHS, PANSS, QLS, GAF and SOFAS, also received benefits in most of the five dimensions of the EQ-5D. Therefore, had we only used a measure of mental health to estimate the benefits of social recovery oriented cognitive-behavioural therapy it could have been that the range of benefits associated with the intervention would have been underestimated.

Implications

Our results suggest that the EQ-5D can discriminate between those with milder and more severe scores, and that the EQ-5D is responsive to improvements in mental health as measured by seven measures of mental health. This suggests that the EQ-5D should be considered for use in future cost-effectiveness studies of mental health interventions. However, as not all of the results were in line with expectations, further research as to the appropriateness of using the EQ-5D in such areas is also warranted.

Garry R. Barton, PhD, **Jo Hodgekins**, BSc, **Miranda Mugford**, DPhil, School of Medicine, Health Policy and Practice, University of East Anglia, Norwich; **Peter B. Jones**, FRCPsych, CAMEO, Cambridgeshire and Peterborough Mental Health Partnership NHS Trust, and Department of Psychiatry, University of Cambridge; **Tim Croudace**, PhD, Department of Psychiatry, University of Cambridge; **David Fowler**, MSc, School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK

Correspondence: Dr Garry Barton, Health Economics Group, School of Medicine, Health Policy and Practice, University of East Anglia, Norwich NR4 7TJ, UK. Email: g.barton@uea.ac.uk

First received 16 Jul 2008, final revision 28 Oct 2008, accepted 20 Jan 2009

Funding

Funding was provided by a trial platform grant from the Medical Research Council (MRC). The MRC had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Acknowledgements

We thank all participants who completed the baseline and 9-month assessment questionnaire, the trial therapists (including Dorothy O'Connor, Annabella Houlden, Neil Harmer, Cas Wright, Mark Wright, Ian Bell, Nick Whitehouse, Patrick Wymbs), Carolyn Crane (Research Nurse), and the UK Mental Health Research Network staff who provided assistance with recruitment and assessment (including Angela Browne, Freya Mellor and Barbara Dickson).

References

- Knapp M, Windmeijer F, Brown J, Kontodimas S, Tziveleakis S, Maria Haro J, et al on behalf of the SOHO Study Group. Cost-utility analysis of treatment with olanzapine compared with other antipsychotic treatments in patients with schizophrenia in the pan-European SOHO study. *Pharmacoeconomics* 2008; **26**: 341–58.
- Barrett B, Byford S, Knapp MJ. Evidence of cost-effective treatments for depression: a systematic review. *J Affect Disord* 2005; **84**: 1–13.
- Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes* (3rd edn). Oxford University Press, 2005.
- Sach TH, Barton GR, Doherty M, Muir K, Jenkinson C, Avery AJ. The relationship between BMI and health related quality of life: comparing the EQ-5D, EuroQol VAS, and SF-6D. *Int J Obes Relat Metab Disord* 2007; **31**: 189–96.
- Knapp M, Thorgrimsen L, Patel A, Spector A, Hallam A, Woods B, et al. Cognitive stimulation therapy for people with dementia: cost-effectiveness analysis. *Br J Psychiatry* 2006; **188**: 574–80.
- McCrone P, Risdale L, Darbishire L, Seed P. Cost-effectiveness of cognitive behavioural therapy, graded exercise and usual care for patients with chronic fatigue in primary care. *Psychol Med* 2004; **34**: 991–9.
- Kennedy TM, Chalder T, McCrone P, Darnley S, Knapp M, Jones RH, et al. Cognitive behavioural therapy in addition to antispasmodic therapy for irritable bowel syndrome in primary care: randomised controlled trial. *Health Technol Assess* 2006; **10**: 1–83.
- Priebe S, Jones G, McCabe R, Briscoe J, Wright D, Slead M, et al. Effectiveness and costs of acute day hospital treatment compared with conventional in-patient care. Randomised controlled trial. *Br J Psychiatry* 2006; **188**: 243–9.
- Dixon L, Hoch JS, Clark R, Bebout R, Drake R, McHugo G, et al. Cost-effectiveness of two vocational rehabilitation programs for persons with severe mental illness. *Psychiatr Serv* 2002; **53**: 1118–24.
- Fenton WS, Hoch JS, Herrell JM, Mosher L, Dixon L. Cost and cost-effectiveness of hospital vs residential crisis care for patients who have serious mental illness. *Arch Gen Psychiatry* 2002; **59**: 357–64.
- Beecham J, Slead M, Knapp M, Chiesa M, Drahorad C. The costs and effectiveness of two psychosocial treatment programmes for personality disorder: a controlled study. *Eur Psychiatry* 2006; **21**: 102–9.
- Byford S, Knapp M, Greenshields J, Ukoumunne OC, Jones V, Thompson S, et al (POMACT Group). Cost-effectiveness of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: a decision-making approach. *Psychol Med* 2003; **33**: 977–86.
- Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Rheumatology* 1997; **36**: 551–9.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res* 2003; **12**: 349–62.
- National Institute for Health and Clinical Excellence. *Guide to the Methods of Technology Appraisal*. NICE, 2008.
- Fowler D, Hodgekins J, Painter M, Reilly T, Crane C, Macmillan I, et al. Cognitive behaviour therapy for improving social recovery in psychosis: a report from the ISREP MRC Trial platform study (Improving Social Recovery in Early Psychosis). *Psychol Med* 2009; Apr (online): 1–10.
- Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull* 1987; **13**: 261–76.
- Beck AT, Steer RA. *Beck Anxiety Inventory*. The Psychological Corporation, 1987.

- 19 Beck AT, Steer RA, Brown GK. *BDI-II Manual*. The Psychological Corporation, 1996.
- 20 Beck AT, Steer RA. *Beck Hopelessness Scale Manual*. The Psychological Corporation, 1988.
- 21 Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry* 1992; **149**: 1148–56.
- 22 Heinrichs DW, Hanlon TE, Carpenter BN. The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophr Bull* 1984; **10**: 388–98.
- 23 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (4th revision, text revision) (DSM-IV-TR)*. APA, 2000.
- 24 Dolan P, Gudex C, Kind P, Williams A. *A social tariff for the EuroQol: results from a UK general population survey (Discussion Paper 138)*. Centre for Health Economics, University of York, 1995.
- 25 Brooks R. EuroQol: the current state of play. *Health Policy* 1996; **37**: 53–72.
- 26 Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997; **35**: 1095–108.
- 27 de Wit GA, Busschbach J, De Charro F. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* 2000; **9**: 109–26.
- 28 Dolan P, Olsen JA. *Distributing Health Care: Economic and Ethical Issues*. Oxford Medical Publications, 2002.
- 29 Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics* 2008; **26**: 769–79.
- 30 Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use* (3rd edn). Oxford University Press, 2003.
- 31 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Criteria for assessing patient based outcome measures for use in clinical trials. *Health Technol Assess* 1998; **14**: 1–74.
- 32 Brazier JE, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999; **3**: 1–164.
- 33 Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care* 2004; **42**: 1125–31.
- 34 Gleitman H. *Psychology* (2nd edn). Norton, 1986.
- 35 Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P. EQ-5D versus SF-6D in an older, chronically ill patient group. *Appl Health Econ Health Policy* 2004; **3**: 91–102.
- 36 Walters SJ, Campbell MJ. The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). *Health Qual Life Outcomes* 2004; **2**: 70.
- 37 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005; **14**: 1523–32.
- 38 Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Muir KR. An assessment of the discriminative ability of the EQ-5D_{index}, SF-6D and EQ VAS, using socio-demographic factors and clinical conditions. *Eur J Health Econ* 2008; **9**: 237–49.
- 39 Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a UK national questionnaire survey. *BMJ* 1998; **316**: 736–41.
- 40 Fayers P, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes* (2nd edn). John Wiley & Sons, 2007.
- 41 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999; **18**: 341–64.
- 42 Palmer S, Davidson K, Tyrer P, Gumley A, Tata P, Norrie J, et al. The cost-effectiveness of cognitive behavior therapy for borderline personality disorder: results from the BOSCO trial. *J Personal Disord* 2006; **20**: 466–81.
- 43 Byford S, Barrett B, Roberts C, Wilkinson P, Dubicka B, Kelvin RG, et al. Cost-effectiveness of selective serotonin reuptake inhibitors and routine specialist care with and without cognitive-behavioural therapy in adolescents with major depression. *Br J Psychiatry* 2007; **191**: 521–7.
- 44 Hakkaart-van Roijen L, van Straten A, Al M, Rutten F, Donker M. Cost-utility of brief psychological treatment for depression and anxiety. *Br J Psychiatry* 2006; **188**: 323–9.
- 45 Konig HH, 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**: 177–87.
- 46 van de Willige G, Wiersma D, Nienhuis F, Jenner J. Changes in quality of life in chronic psychiatric patients: a comparison between EuroQol (EQ-5D) and WHOQoL. *Qual Life Res* 2005; **14**: 441–51.
- 47 Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999; **38**: 870–7.
- 48 Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* 2005; **60**: 1571–82.
- 49 Russell AS, Conner-Spady B, Mintz A, Maksymowich WP. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *J Rheumatol* 2003; **30**: 941–7.
- 50 Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multi-attribute and single attribute utility functions for the Health Utilities Index Mark 3 system. *Med Care* 2002; **40**: 113–28.
- 51 Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; **21**: 271–92.
- 52 Ware JE, Sherbourne C. The MOS 36 item short-form health survey. I. Conceptual framework and item selection. *Med Care* 1992; **30**: 473–83.
- 53 Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? *Eur J Health Econ* 2005; **6**: 253–60.
- 54 Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Econ* 2006; **15**: 1229–36.