This is an author produced version of Usefulness of EQ-5D in Assessing Health Status in Primary Care Patients with Major Depressive Disorder.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/1037/

Article:
Sapin, C., Fantino, B., Nowicki, M.L. et al. (1 more author) (2004) Usefulness of EQ-5D in Assessing Health Status in Primary Care Patients with Major Depressive Disorder. Health and Quality of Life Outcomes. ISSN 1477-7525
Usefulness of EQ-5D in Assessing Health Status in Primary Care Patients with Major Depressive Disorder

Christophe Sapin*1, Bruno Fantino2, Marie-Laure Nowicki3 and Paul Kind4

Abstract

Objectives: Major depressive disorder (MDD) is a prevalent psychiatric disorder associated with impaired patient functioning and reductions in health-related quality of life (HRQL). The present study describes the impact of MDD on patients' HRQL and examines preference-based health state differences by patient features and clinical characteristics.

Methods: 95 French primary care practitioners recruited 250 patients with a DSM-IV diagnosis of MDD for inclusion in an eight-week follow-up cohort. Patient assessments included the Montgomery Asberg Depression Rating Scale (MADRS), the Clinical Global Impression of Severity (CGI), the Short Form-36 Item scale (SF-36), the Quality of Life Depression Scale (QLDS) and the EuroQoL (EQ-5D).

Results: The mean EQ-5D utility at baseline was 0.33, and 8% of patients rated their health state as worse than death. There were no statistically significant differences in utilities by demographic features. Significant differences were found in mean utilities by level of disease severity assessed by CGI. The different clinical response profiles, assessed by MADRS, were also revealed by EQ-5D at endpoint: 0.85 for responders remitters, 0.72 for responders non-remitter, and 0.58 for non-responders. Even if HRQL and EQ-5D were moderately correlated, they shared only 40% of variance between baseline and endpoint.

Conclusions: Self-reported patient valuations for depression are important patient-reported outcomes for cost-effectiveness evaluations of new antidepressant compounds and help in further understanding patient compliance with antidepressant treatment.

Introduction

Major depressive disorder (MDD) is common in primary care patients [1] with a lifetime prevalence rate in the French population of 10–25% in women and 5–12% in men [2]. Depression is associated with marked decreases in functioning, well being and health-related quality of life (HRQL) [3,4], and an increases in disability days [5], use of health services and overall societal costs [6]. Antidepressant treatments are effective in reducing depression...
severity [7,8] and in increasing patient functioning and HRQL [9,10].

The Washington Panel on Cost-Effectiveness in Health and Medicine recommended the use of HRQL in the evaluation of health care interventions [11]. For this purpose, HRQL measurement needs to express patient health status on a scale where perfect health and death are valued 1 and 0 respectively. When such quality of life data are combined with corresponding data on the quantity of life, then the consequences of treatment are measured in units of Quality-Adjusted Life Years (QALYs) [12]. Where QALYs are calculated for social decision making purposes then the HRQL measures used to make the quality adjustment should be based on the preferences of the population as a whole. Such social preferences are only available for a limited number of HRQL measures and for a limited number of countries. EQ-5D is one such measure that has been calibrated in this way.

Studies of physical illnesses have suggested that patient’s values for their own health state affect decisions concerning treatment and its outcomes [13-15]. Several studies focused on establishing utility scores for a variety of health states in various mental illnesses, including schizophrenia [16], depression in primary care [17], temporary states of depression [18,19] and treatment-related side effects [20,21]. Health states in depression have been characterised by the presence or absence of symptoms, and depressed patients are usually categorised as responders and remitters using classical rating scales [22]. As responders sometimes present residual depressive symptoms, we classify patients as "Responder remitters", "Responders non-remitter" and "Non-responders".

The objectives of this paper are to describe the impact on HRQL of patients with MDD treated in a primary care setting, and to examine variations in terms of patients’ demographic and clinical characteristics.

Material and methods

Design and patient sample

This national, multicentre, prospective, non-comparative cohort study was designed so as to reproduce the guidelines for management of depression in primary care. The scheduled follow-up period was two months, with assessments at baseline (D0), four weeks later (D28) and eight weeks later (D56).

The patients included in this study were recruited from an outpatient population, aged 18 and older, who consulted general practitioners for a new episode of MDD according to the DSM-IV [23], and who were not treated with any antidepressant before inclusion. Patients whose symptomatology suggested schizophrenia or other psychotic symptoms, according to DSM-IV, were not included in this study. According to their experience and daily practice, general practitioners initiated an antidepressant treatment at baseline.

Data collection

Patients’ characteristics

Patient profiles were created at baseline by recording age, gender, lifestyle, place of residence, socio-professional category and current professional status.

Clinical measures

Physicians assessed the severity of depressive symptoms using the Montgomery-Asberg Depression Rating Scale (MADRS) [24] and the Clinical Global Impression of Severity (CGI-S) scale. The CGI was rated by physicians on a seven-point Likert scale ranging from 1 = "Normal, not ill at all" to 7 = "Among the most ill patients".

Qualitative outcomes derived from rating scales, like response to treatment or remission, are usually used in both clinical trials and economic evaluations of new antidepressant agents [22]. Using MADRS scores at D56, patients were classified into two groups: those that had scores lower or equal to 12 were considered as "Remitters", the others were considered as "Non-remitters". Patients who had a decrease of at least 50% in relation to baseline score were considered as "Responder", whereas the others were "Non-responders". These two patient groupings led to the creation of three mutually exclusive groups: "Responder remitters", "Responders non-remitter" and "Non-responders".

Patient Reported Outcomes

The outcome measures used in this study were the 36-item Short-Form Health Survey (SF-36), the Quality of Life in Depression Scale (QLDS) and the EQ-5D.

The SF-36 is a generic HRQL measure consisting of eight dimensions assessing physical functioning (PF), role limitations due to physical problems (RP), bodily pain (BP), general health (GH), vitality (VT), mental health (MH), role limitations due to emotional problems (RE) and social functioning (SF) [25]. Two summary scores also assess both physical (PCS) and mental (MCS) facets [26]. All scale scores range from 0 (the worst HRQL) to 100 (the best HRQL).

The QLDS is a 34-item depression-specific HRQL instrument that assesses the ability and capacity of individuals to satisfy their daily needs [27,28]. Each item is answered by Yes or No. An overall HRQL score is obtained by summing the 34 items. The results range from 0 (the highest HRQL) to 34 (the lowest HRQL).
EQ-5D is a generic measure of HRQL in which health status is defined in terms of 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression [29]. Each dimension has three qualifying levels of response roughly corresponding to 'no problems', 'some difficulties/problems', and 'extreme difficulties'. EQ-5D defines a total of 243 unique health states. The importance of each of these states can be determined in a number of different ways. For the purpose of cost-utility analysis and other situations where the consequences of treatment are measured in terms of QALYs, these weights are typically established using utility measurement techniques such as Standard Gamble or Time Trade-Off (TTO) [12]. For the purposes of this present study, TTO weights elicited from a large national survey of the UK population were used [30]. Information collected using EQ-5D can be reported in terms of its individual dimensions and as a single index score (EQ-5D

Data analysis
Continuous variables were expressed by means and standard deviations, whereas categorical data were presented using frequency and percentage. The scales were scored using scoring algorithms described by the scale designers. Student's t-tests, ANOVA, Mann-Whitney, or Kruskal-Wallis tests were performed when appropriate to compare mean scores across subgroups. Regression analyses were used to examine the relationships between differences in the utility-weighted EQ-5D

Results
Sample characteristics
Ninety-five physicians enrolled 250 patients between May and November 2002. Patient age ranged from 18 to 92 years, with a mean of 44.2 ± 14.1 years (mean ± standard deviation). The sex ratio (males/females) was 0.4. The mean MADRS score was 32.7 ± 7.7, ranging from 13 to 53. This high level of severity was also revealed by the CGI: about 85% of patients were rated "markedly ill" or more severely. The demographic and clinical characteristics of the sample are reported in Table 1.

Among the 250 included patients, 24 were lost to follow-up (9.6%). Their sociodemographics and clinical charac-
teristics were not significantly different from those of the 226 completers, so that all subsequent analyses were performed on the completers sub-sample.

Impact of MDD on HRQL
At baseline, the mean QLDS score was 20.8 ± 5.8, ranging from 5 to 31. The mean SF-36 dimension scores for the total sample were: PF 69.0 ± 24.5, RP 22.4 ± 30.7, BP 52.0 ± 23.5, GH 38.3 ± 17.3, VT 22.2 ± 13.1, MH 24.5 ± 12.1, RE 9.1 ± 21.3 and SF 30.2 ± 17.1. The mean SF-36 summary scores were PCS 43.6 ± 9.2 and MCS 21.1 ± 6.6. The mean EQ-5D$_{ST}$ score was 0.33 ± 0.25, ranging between -0.59 and 0.85. It was noteworthy that 8% of the study population had an EQ-5D$_{ST}$ score of worse than death, i.e. less than zero.

During follow-up, all the dimensions rated improved (Table 2). The mean EQ-5D$_{ST}$ scores were 0.68 ± 0.24 (range: [-0.11; 1.00]) and 0.78 ± 0.21 (range: [-0.08; 1.00]) at weeks 4 and 8, respectively.

Comparison of EQ-5D$_{ST}$ by demographic and clinical features
No significant differences were found in EQ-5D$_{ST}$ by demographics characteristics (Table 3): men and women reported the same preference-based score at baseline (0.32 ± 0.22 vs. 0.32 ± 0.26, respectively) and their scores increased in a similar manner during follow-up. Younger patients reported higher utility scores than older patients at baseline, day 28 and day 56, although this pattern was not statistically significant.

Significant differences in EQ-5D$_{ST}$ were found by disease severity level assessed by CGI-S, with more severe patients having lower weighted index scores. At baseline, a mean difference of 0.12 was observed between "slightly/moderately ill" and "markedly ill" patients (p < 0.05), and 0.18 between "markedly ill" and "seriously ill" patients (p < 0.001). At the end of the follow-up, a mean difference of 0.12 was observed between patients with "first signs of illness" and "slightly/moderately ill" patients (p < 0.001). "Slightly/moderately ill" and "markedly ill" patients had EQ-5D$_{ST}$ scores that differed 0.30 on average (p < 0.001). A mean difference of 0.14 between "markedly ill" patients and "seriously ill" patients was found (p < 0.05).

Comparison of EQ-5D$_{ST}$ by clinical response
Clinical response, defined by MADRS scores, revealed statistically significant differences between mean EQ-5D$_{ST}$ scores at baseline (p < 0.01), D28 (p < 0.001) and D56 (p < 0.001) (Table 4).

At baseline, an overall significant difference was found in comparing the three groups, with a mean difference of 0.14 observed between "Responder remitters" and "Responder non-remitters" (p < 0.01). During the study period, EQ-5D$_{ST}$ scores increased in all groups of clinical response. At the end of the follow-up, a statistically significant mean difference of 0.14 was observed between

---

Table 2: EQ-5D dimension scores at baseline, D28 and D56

<table>
<thead>
<tr>
<th>Mobility</th>
<th>Baseline</th>
<th>D28</th>
<th>D56</th>
</tr>
</thead>
<tbody>
<tr>
<td>No difficulties</td>
<td>166</td>
<td>192</td>
<td>207</td>
</tr>
<tr>
<td>Some difficulties</td>
<td>58</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>Extreme difficulties</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Self-care
- No difficulties: 186, N%: 82.3, 214, N%: 58.9, 172, N%: 76.1
- Some difficulties: 58, N%: 25.6, 12, N%: 4.4
- Extreme difficulties: 2, N%: 0.9, 0, N%: 0.0

Usual activities
- No difficulties: 56, N%: 24.8, 133, N%: 50.4, 172, N%: 76.1
- Some difficulties: 150, N%: 66.4, 92, N%: 40.7, 53, N%: 23.5
- Extreme difficulties: 20, N%: 8.8, 1, N%: 0.4

Pain/Discomfort
- No difficulties: 54, N%: 23.9, 114, N%: 50.4, 135, N%: 59.8
- Some difficulties: 153, N%: 67.7, 108, N%: 47.8, 90, N%: 39.8
- Extreme difficulties: 19, N%: 8.4, 4, N%: 1.8

Anxiety/Depression
- No difficulties: 2, N%: 0.9, 27, N%: 11.9, 72, N%: 31.9
- Some difficulties: 48, N%: 21.2, 155, N%: 68.6, 133, N%: 58.8
- Extreme difficulties: 176, N%: 77.9, 44, N%: 19.5, 21, N%: 9.3

---
"Responder remitters" and "Responders non-remitter" (p < 0.001). "Responders non-remitter" and "Non-responders" had EQ-5D ST scores that significantly differed by 0.14 on average (p < 0.05).

Comparison of Patient Reported Outcomes
At each visit, EQ-5D ST scores were compared with SF-36 dimension, SF-36 summary and QLDS scores (Table 5) by computing correlation coefficients. The correlation between EQ-5D ST score and the Mental Health dimension of the SF-36 was the highest observed, whatever the assessment (DO: r = 0.49; D28: r = 0.56; D56: r = 0.63). At baseline, Pearson correlation coefficients were always greater than 0.30, except for the role-physical and role-emotional dimensions.

The QLDS was significantly correlated with the EQ-5D ST scores, ranging from -0.43 at baseline to -0.68 at the end of the follow-up period.

Multivariate analysis
An ordinary least-square regression analysis to predict EQ-5D ST using demographic features, clinical and HRQL evolution only explained 40% of the variance in the weighted index scores. The statistically significant predictors in the regression model were differences in Physical Functioning, Bodily Pain, General Health and Mental Health (Table 6).

Discussion
This study evaluated the usefulness of EQ-5D in assessing health status of primary care patients with major depressive disorder.

The sampling of our study is representative of the primary care depressed population in France [2]. 8% of the patients rated their health state as worse than death. This result is not surprising given the relationship between depression and suicide [31,32]. Despite different approaches to measuring health state utilities using standard gamble, time trade-off or rating scales, the findings of

---

**Table 3: Differences on utility score by demographic and clinical characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>D28</th>
<th>D56</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>p</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.32 ± 0.22</td>
<td>0.94</td>
<td>0.66 ± 0.24</td>
</tr>
<tr>
<td>Female</td>
<td>0.32 ± 0.26</td>
<td>0.76</td>
<td>0.68 ± 0.25</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤30 years</td>
<td>0.35 ± 0.23</td>
<td>0.76</td>
<td>0.76 ± 0.24</td>
</tr>
<tr>
<td>31–65 years</td>
<td>0.32 ± 0.24</td>
<td>0.76</td>
<td>0.66 ± 0.25</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td>0.30 ± 0.38</td>
<td>0.76</td>
<td>0.68 ± 0.17</td>
</tr>
<tr>
<td>Place of residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban area</td>
<td>0.31 ± 0.25</td>
<td>0.76</td>
<td>0.76 ± 0.24</td>
</tr>
<tr>
<td>Rural area</td>
<td>0.34 ± 0.25</td>
<td>0.76</td>
<td>0.66 ± 0.25</td>
</tr>
<tr>
<td>CGI-Severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First signs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.80 ± 0.14</td>
<td>&lt;0.001</td>
<td>0.86 ± 0.13</td>
</tr>
<tr>
<td>Slightly/moderately ill</td>
<td>0.45 ± 0.22</td>
<td>&lt;0.001</td>
<td>0.66 ± 0.23</td>
</tr>
<tr>
<td>Markedly ill</td>
<td>0.33 ± 0.24</td>
<td>&lt;0.001</td>
<td>0.40 ± 0.26</td>
</tr>
<tr>
<td>Seriously ill</td>
<td>0.15 ± 0.21</td>
<td>&lt;0.001</td>
<td>0.14 ± 0.14</td>
</tr>
</tbody>
</table>

**Table 4: Utility scores and clinical response during the study period**

<table>
<thead>
<tr>
<th>Clinical response</th>
<th>Baseline</th>
<th>D28</th>
<th>D56</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>p</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Responder remitters</td>
<td>0.35 ± 0.24</td>
<td>&lt;0.01</td>
<td>0.76 ± 0.18</td>
</tr>
<tr>
<td>Responders non-remitter</td>
<td>0.21 ± 0.25</td>
<td>&lt;0.001</td>
<td>0.54 ± 0.26</td>
</tr>
<tr>
<td>Non-responders</td>
<td>0.30 ± 0.27</td>
<td>&lt;0.001</td>
<td>0.54 ± 0.30</td>
</tr>
</tbody>
</table>
our study agree with those previously reported: the base-
line mean utility of an untreated depression was 0.33,
compared to 0.30 for Revicki [21] and 0.32 for Bennett
[33]. Patient-rated EQ-5D ST scores after the eight-week fol-
low-up period was 0.78, which is comparable to utilities
reported in other studies (0.79 [33]; 0.74 [21]; 0.76 [34];
0.70 [35]). The main interest is that EQ-5D ST values are
easy to collect in large sample surveys due to the brevity of
EQ-5D classification system with its 5 dimensions and 3
levels.

No differences in EQ-5D ST utilities were observed by
demographic characteristics, which is comparable to pre-
vious results in depressed patients [21,34,35]. More
severely depressed patients reported utilities that were
0.30 points lower than less severely depressed patients at
baseline. Several researchers have suggested that dif-
ferences in utility greater than 0.05 are clinically important
[12,36]. These findings may reflect clinically important
differences.

As demonstrated in previous studies [21,37], we found
that the EQ-5D ST score and other HRQL measures shared
only about 40% of variance. Utilities measure a patient’s
preference for their health state, while HRQL scales assess
the patient’s report of their functioning and well-being. 
Although these two concepts are related they are not iden-
tical [38], and measuring both may lead to a better under-
standing of reasons for non-compliance to treatment
regimens.

There are several limitations that need to be considered
when interpreting the results of this study. First, the study
does not take into account the antidepressant prescribed
or their side effects, which may influence patients’ ratings
[21,39]. Second, the concomitant impact of depression and
chronic medical conditions could not be examined in
this sample. It is likely that the health state utilities of
patients with depression, in addition to a chronic medical
disease would be significantly reduced [17]. Lastly, a lim-
itation of the analysis presented in this study relates to the
source of the utility weights used to compute the EQ-5D ST.
Given that this was a national study conducted in France
it may have been better to use social preference values
based on the French population. Unfortunately, at the
time of writing these values were not available for EQ-

---

Table 5: Association between utility score and HRQL

<table>
<thead>
<tr>
<th>Short Form-36 Item</th>
<th>Baseline Correlation</th>
<th>D28 Correlation</th>
<th>D56 Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Functioning</td>
<td>0.39***</td>
<td>0.42***</td>
<td>0.45***</td>
</tr>
<tr>
<td>Role-Physical Limitations</td>
<td>0.28***</td>
<td>0.51***</td>
<td>0.47***</td>
</tr>
<tr>
<td>Bodily Pain</td>
<td>0.37***</td>
<td>0.53***</td>
<td>0.53***</td>
</tr>
<tr>
<td>General Health</td>
<td>0.47***</td>
<td>0.54***</td>
<td>0.58***</td>
</tr>
<tr>
<td>Vitality</td>
<td>0.45***</td>
<td>0.54***</td>
<td>0.59***</td>
</tr>
<tr>
<td>Mental Health</td>
<td>0.49***</td>
<td>0.55***</td>
<td>0.63***</td>
</tr>
<tr>
<td>Role-Emotional Limitations</td>
<td>0.26***</td>
<td>0.46***</td>
<td>0.49***</td>
</tr>
<tr>
<td>Social Functioning</td>
<td>0.44***</td>
<td>0.54***</td>
<td>0.59***</td>
</tr>
<tr>
<td>Physical Composite Summary</td>
<td>0.42***</td>
<td>0.51***</td>
<td>0.50***</td>
</tr>
<tr>
<td>Mental Composite Summary</td>
<td>0.34***</td>
<td>0.49***</td>
<td>0.58***</td>
</tr>
<tr>
<td>Quality of Life for Depression Scale</td>
<td>-0.43***</td>
<td>-0.59***</td>
<td>-0.68***</td>
</tr>
</tbody>
</table>

*** p < 0.001

Table 6: Contributors of the difference in EQ-5D ST during the study period

<table>
<thead>
<tr>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>p-value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1971</td>
<td>0.0291</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Physical Functioning</td>
<td>0.0019</td>
<td>0.0007</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Bodily Pain</td>
<td>0.0016</td>
<td>0.0007</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>General Health</td>
<td>0.0030</td>
<td>0.0009</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Mental Health</td>
<td>0.0044</td>
<td>0.0009</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
5D-ST. Weights were therefore adopted from a major UK study that provided the most robust technical estimates widely used in the evaluation of EQ-5D-ST in countries that lack their own national reference data.

Utility scores are needed for calculating QALYs, which are used as indicators of effectiveness or outcome in economic evaluations [35, 36, 40]. It is debatable whether or not patient or general population utilities should be used in cost-effectiveness studies [40]. Nevertheless, patients with experience in the disease may be the best providers of health state preference data. Cost-effectiveness studies are required to help clinicians and health care decision-makers in determining the impact of new antidepressants on both patient outcomes and medical or overall societal costs. Understanding patient preferences for depression outcomes is important for economic evaluations of new antidepressants, as well as for understanding patient behaviour and compliance to antidepressant regimens. Such a measure can be applied to cost-utility analyses either within clinical decision modelling studies or within prospective, randomised clinical trials and offers additional scope for the analysis and reporting of data derived from clinical trials of new compounds.

Acknowledgment
Funding for this study was provided by H. Lundbeck A/S.

References

